



**SIRALI ÖLÇEK SKORLARI İÇİN DEĞERLENDİRİCİLER-ARASI
GÜVENİLİRLİK ÖLÇÜM METOTLARININ KARŞILAŞTIRILMASI**

Nilden TUTALAR

**YÜKSEK LİSANS TEZİ
İSTATİSTİK ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

HAZİRAN 2015

Nilden TUTALAR tarafından hazırlanan “SIRALI ÖLÇEKLERDE DEĞERLENDİRİCİLER-ARASI GÜVENİRLİK ÖLÇÜM YÖNTEMLERİNİN KARŞILAŞTIRILMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi İstatistik Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Bülent ÇELİK

İstatistik Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Başkan: Prof. Dr. Birdal ŞENOĞLU

İstatistik Anabilim Dalı, Ankara Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Üye: Doç. Dr. H. Hasan ÖRKÇÜ

İstatistik Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Tez Savunma Tarihi: 10/06/2015

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....
Prof. Dr. Şeref SAĞIROĞLU

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Nilden TUTALAR

17.06.2015

SIRALI ÖLÇEKLERDE DEĞERLENDİRİCİLER-ARASI GÜVENİRLİK ÖLÇÜM YÖNTEMLERİNİN KARŞILAŞTIRILMASI

(Yüksek Lisans Tezi)

Nilden TUTALAR

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2015

ÖZET

Tıp, sosyal bilimler, eğitim bilimleri, dil, vb. pek çok alanda son yirmi yıldır sıklıkla bahsi geçen bir analiz olan değerlendiriciler arası güvenirlik analizi çeşitli ölçek tipleri için geliştirilen birçok uyum katsayısıyla gerçekleştirilebilmektedir. Sınıflama, sıralama, eşit aralıklı ve oranlama ölçme düzeyleri için geliştirilen mevcut uyum istatistikleri halen incelenmekte ve geliştirilmektedir. İki veya daha fazla değerlendiricinin sıralı ölçek üzerinde her deneği ayrı ayrı sınıflandırdığı durumlarla sıklıkla karşılaşılmaktadır. Birçok sıralı ölçek, bir deneğin belli bir duruma sahip olup olmadığını ortaya çıkarmada oldukça öznelir. Belirli bir sıralamaya sahip kategorilerden oluşan bir değerlendiriciler arası güvenirlik deneyinde iki veya daha fazla değerlendiricinin sübjektif veya belirli bir prosedürü takiben yapmış olduğu skorlamaların uyumunu ölçmek için literatürde çeşitli uyum istatistikleri önerilmiştir. Bunlar iki ve ikiden fazla değerlendirici için olmak üzere gruplandırılabilir. İki değerlendirici için, Cohen Ağırlıklandırılmış κ katsayısı, Scott Ağırlıklandırılmış π katsayısı, Brennan-Prediger uyum katsayısı, Krippendorff α katsayısı ve Gwet AC₂ uyum katsayısı sıklıkla kullanılırken, ikiden fazla değerlendirici için ise Conger Ağırlıklandırılmış κ katsayısı, Fleiss Genelleştirilmiş κ katsayısı, Brennan-Prediger uyum katsayısı, Krippendorff α katsayısı ve Gwet AC₂ uyum katsayısı kullanılmaktadır. Bu tezin amacı, sıralı ölçekli skorlamalara sahip durumlarda değerlendiriciler arasındaki uyumu ölçen uyum katsayılarının karşılaştırılması ve bu uyum katsayılarının gücünün değerlendirilmesidir. Gerçek veriler kullanılarak yapılan hesaplamalar sonucunda, Gwet AC₂ katsayısının diğer uyum katsayıları içerisinde en büyük uyum değerini verdiği tespit edilmiştir. Uyum gücünün yorumlanmasında ise literatürdeki mevcut Landis ve Koch, Fleiss ve Altman kıyaslama ölçekleri ele alınarak Gwet (2014) alternatif kıyaslama yöntemi incelenmiştir. Herhangi bir uyum katsayısını direk olarak kıyaslama ölçekleriyle yorumlamaktan ziyade Gwet (2014) kıyaslama yöntemini kullanarak yorumlamanın istatistiksel olarak daha güvenilir olduğu tespit edilmiştir.

Bilim Kodu : 205.1.066

Anahtar Kelimeler : Değerlendiriciler-arası güvenirlik; sıralı skorlar; uyum; uyum katsayısı; ölçüm; gözlemciler-arası uyum

Sayfa Adedi : 125

Danışman : Doç. Dr. Bülent ÇELİK

COMPARISON OF INTER-RATER RELIABILITY MEASUREMENT METHODS IN ORDINAL SCALES

(M. Sc. Thesis)

Nilden TUTALAR

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June, 2015

ABSTRACT

Inter-rater reliability analysis that often mentioned in medical, social sciences, pedagogics, and linguistics e.t.c in the past few decades can be performed with several agreement coefficients that developed for various scale type. The current agreement coefficients that developed for nominal, ordinal, interval and ratio scales have been still researched and improved. The cases in which two or more raters separately classify each subject in a sample on an ordinal scale are frequently encountered. Many ordinal scales are quite subjective in determining about whether a subject has a certain condition or not. Many ordinal scales are quite subjective, such as the scale for diagnoses about whether a subject has a certain condition. In an inter-rater reliability experiment that consist of categories that having a particular order there are several agreement coefficients are proposed in the literature for measuring scores that are scored by two or more raters following a particular or subjective procedure. These are grouped in terms of two or more than two raters. For 2 raters case these are Cohen's Weighted κ (1968), Scott's Weighted π (1955), Brennan-Prediger B-P (1981), Krippendorff's α (1970, 1978, 2004) and Gwet's AC₂ (2008). For more than 2 raters case these are Conger's Generalized Kappa (1980), Brennan-Prediger B-P (1981), Fleiss' Generalized κ (1971), Krippendorff's α (1970, 1978, 2004) and Gwet's AC₂ (2008). This dissertation concerned with comparing agreement coefficients that measure agreement among raters in an ordinal scaled inter-rater reliability experiment and evaluating these coefficients' strength. As a result of calculations which done by using real data, it has been seen that Gwet AC₂ yields the higher agreement value than the other coefficients. On the interpretation of the strength of agreement Gwet (2014) alternative benchmarking method is analyzed by comparing existing benchmarking scales such as Landis ve Koch, Fleiss and Altman in literature. It was determined that interpreting the strength of agreement by using Gwet benchmarking method is more reliable rather than interpreting it by directly comparing agreement coefficients with existing benchmarking scales.

Science Code : 205.1.066

Key Words : inter-rater reliability; ordinal scaled scores; agreement; agreement coefficients; measurement; inter-observer reliability

Page Number : 125

Supervisor : Assoc.Prof.Dr. Bülent ÇELİK

TEŞEKKÜR

Bu çalışmanın her aşamasında ilgi ve desteği ile beni yönlendiren tez danışmanım Doç. Dr. Bülent ÇELİK'e, tezimle ilgili yapıcı eleştiri ve önerileriyle beni yönlendiren hocalarım Prof. Dr. Birdal ŞENOĞLU ve Doç. Dr. H. Hasan ÖRKÇÜ'ye; bu tezin konusunda uzman olan ve takıldığım her noktada bana ABD'den mail ile desteğini esirgemeyen bilim adamı sayın Ph.D. Kilem Li GWET'e; bugünlere gelmemi sağlayan ve benden desteğini hiçbir zaman esirgememiş kıymetli babam Hüseyin TUTALAR ve değerli annem Nermin TUTALAR'a; tezin yazım aşamasında ve IBS EMR 2015 Eastern Mediterranean Region International Biometric Society konferansı sırasında desteklerini esirgemeyen arkadaşlarım Burçin ÖNER ve Esra ARSLANOĞLU'na; yabancı kaynaklar konusunda çevirilerimi kontrol edip düzelten değerli hocalarım Ar. Grv. Zeynep Nur TOKER ve Ar. Grv. Mustafa YAMAN ve ingilizce öğretmeni Neslihan SÖZEN'e; bu tez konusunu oluşturan teknikleri dilbilim alanında gerçek bir uygulama üzerinde uygulamamı sağlayıp bana bu tezi yazmamda ilham veren Ar. Grv. Anıl ALBAĞLAR'a; yüksek lisansa girişimden itibaren benden hiçbir desteğini ve güler yüzünü esirgemeyen, her zaman yardımcı olmaya çalışan Gazi Üniversitesi Fen Bilimleri Enstitüsü'nde İstatistik bölümündeki öğrenci işleri sorumlusu Hülya AYDIN'a ve yardımı dokunan tüm enstitü çalışanlarına; karakteriyle, hayata bakış açısıyla kendisini tanımaktan son derece mutluluk duyduğum ve bana her anımda destek olan dostum İlknur CİHANGİR'e; sosyal paylaşım sitesi twitter hesabımdan bana destek olan ve sadece istatistik bilimi ile kalmayıp ufku açan başta astronomi, fizik, biyoloji, tıp, evrimsel biyoloji, istatistik, matematik, ekonometri, bilgisayar ve yazılım mühendisliği alanındaki tüm bilim adamlarına ve bilim aşıklarına; son olarak tezim esnasında ve tezimi bitirirken destekleri kadar aynı zamanda her türlü engel ve sıkıntıyı da tezat bir şekilde yaşatan kişilere de sorun ve problem çözme yeteneğimi daha çok geliştirdikleri için teşekkürlerimi sunuyorum.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
SİMGELER VE KISALTMALAR.....	xvii
1. GİRİŞ	1
2. SIRALI ÖLÇEK SKORLARI İÇİN DEĞERLENDİRİCİLER-ARASI GÜVENİLİRLİK ÖLÇÜMÜ YÖNTEMLERİ	5
2.1. Sıralı Ölçek Kavramı	5
2.2. Uyum Kavramı	6
2.2.1. Şansa bağlı uyum kavramı(pe).....	7
2.2.2. Kısmi uyum kavramı.....	7
2.3. Ağırlıklandırma Kavramı.....	8
2.3.1. Kuadratik ağırlıklar	8
2.3.2. Basit sıralı ağırlıklar	9
2.4. İki Değerlendiricili Durum için Uyum Katsayıları.....	10
2.4.1. Cohen'in ağırlıklandırılmış kappa katsayısı	10
2.4.2. Ağırlıklandırılmış Scott'un π katsayısı	14
2.4.3. Ağırlıklandırılmış Brennan-Prediger BP katsayısı	16
2.4.4. Krippendorff'un α katsayısı	18
2.4.5. Gwet AC ₂ katsayısı	19
2.5. Üç veya Daha Fazla Değerlendiricili Durum İçin Uyum Katsayıları.....	21

	Sayfa
2.5.1. Conger'in kappası	23
2.5.2. Brennan-Prediger katsayısı	25
2.5.3. Fleiss'in Kappa katsayısı	25
2.5.4. Krippendorff'un α katsayısı	26
2.5.5. Gwet AC ₂ katsayısı	28
3. DEĞERLENDİRİCİLER-ARASI GÜVENİLİRLİK KATSAYISINI KİYASLAMA YÖNTEMLERİ	31
3.1. Kıyaslama Kavramı ve Uyumun Büyüklüğü	31
3.2. Deneysel Koşulların Etkisi	33
3.3. Literatürdeki kıyaslama yöntemleri	34
3.3.1. Landis ve Koch kıyaslama ölçeği	34
3.3.2. Fleiss kıyaslama ölçeği	34
3.3.3. Altman kıyaslama ölçeği	35
3.3.4. Gwet'in kıyaslama modeli	36
4. BULGULAR	47
4.1. İki Değerlendiricili Durum	47
4.1.1. Üç kategorili durum	47
4.1.2. Dört kategorili durum	53
4.1.3. Beş kategorili durum	59
4.1.4. Altı kategorili durum	64
4.1.5. Yedi kategorili durum	69
4.2. Üç ve Daha Fazla Değerlendiricili Durumlar	73
4.2.1. Üç değerlendirici	73
4.2.2. Dört değerlendirici	77

	Sayfa
4.2.3. Beş değerlendirici	81
4.2.4. Yedi değerlendirici.....	86
5. TARTIŞMA VE SONUÇ	91
KAYNAKLAR	95
EKLER.....	99
EK-1. Kuadratik (Karesel) ve basit sıralı ağırlıkların hesaplanmasına ilişkin örnek bir uygulama.....	100
EK-2. Kappa: karesi alınmış öklid uzaklıklarının bir fonksiyonu	103
EK-3. İki değerlendiricili durum için örnek uygulama örneği	105
EK-4. Üç ve daha fazla değerlendiricili durum için uyum hesaplaması örnek uygulama	113
EK-5. Landis ve Koch (1977b) verisi	123
EK-6. Stacklefish verisi	124
ÖZGEÇMİŞ	125

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Landis ve Koch (1977a) kappa için belirlediği kıyaslama ölçeği.....	34
Çizelge 3.2. Fleiss (1981) kappa için belirlediği kıyaslama ölçeği	35
Çizelge 3.3. Altman(1991) kappa için belirlediği kıyaslama ölçeği.....	35
Çizelge 3.4. Landis ve Koch (1977b) verisine ilişkin hesaplanan değerlendiriciler-arası güvenirlilik katsayıları.....	39
Çizelge 3.5. Landis ve Koch (1977b) verisine ilişkin hesaplanan değerlendiriciler-arası güvenirlilik katsayılarının Landis&Koch, Fleiss ve Altman ölçeklerine göre kıyaslama değerleri	40
Çizelge 3.6. Uyum katsayılarına göre Landis&Koch kıyaslama aralıklarına ait aitlik olasılıkları.....	40
Çizelge 3.7. Uyum katsayılarına göre Landis&Koch kıyaslama aralıklarına ait aitlik kümülatif olasılıkları	40
Çizelge 3.8. Gwet'in yeni yöntemi altında kıyaslama	41
Çizelge 3.9. Uyum katsayılarına göre Fleiss kıyaslama aralıklarına ait aitlik olasılıkları	42
Çizelge 3.10. Uyum katsayılarına göre Fleiss kıyaslama aralıklarına ait aitlik kümülatif olasılıkları.....	42
Çizelge 3.11. Gwet'in yeni yöntemi altında kıyaslama	42
Çizelge 3.12. Uyum katsayılarına göre Altman kıyaslama aralıklarına ait aitlik olasılıkları.....	43
Çizelge 3.13. Uyum katsayılarına göre Altman kıyaslama aralıklarına ait aitlik kümülatif olasılıkları	43
Çizelge 3.14. Gwet'in yeni yöntemi altında kıyaslama	43
Çizelge 3.15. Kıyaslama Ölçekleri Bazında Karşılaştırma Çizelgesi.....	44
Çizelge 4.1. Deneklerin değerlendirici ve kategori bazında dağılımı.....	48
Çizelge 4.2. Çizelge 4.1'in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar	48
Çizelge 4.3. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	48

Çizelge	Sayfa
Çizelge 4.4. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	50
Çizelge 4.5. Deneklerin değerlendirici ve kategori bazında dağılımı.....	54
Çizelge 4.6. Çizelge 4.1'in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar	54
Çizelge 4.7. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	54
Çizelge 4.8. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	56
Çizelge 4.9. Rahim serviksde karsinoma in situnun klasifikasyon verisi.....	59
Çizelge 4.10. Çizelge 4.9'in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar ..	60
Çizelge 4.11. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	60
Çizelge 4.12. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	61
Çizelge 4.13. Beyin enfarktüsü teşhisi verisi.....	65
Çizelge 4.14. Çizelge 4.13'ün analizi için kullanılan kuadratik ve basit sıralı ağırlıklar	66
Çizelge 4.15. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	66
Çizelge 4.16. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	67
Çizelge 4.17. Engellilik derecesi sınıflandırması ham verisi.....	69
Çizelge 4.18. Çizelge 4.17'nin analizi için kullanılan kuadratik ve basit sıralı ağırlıklar	70
Çizelge 4.19. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	70
Çizelge 4.20. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	71
Çizelge 4.21. Rahim serviksde karsinoma in situnun klasifikasyon verisi.....	73
Çizelge 4.22. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	74

Çizelge	Sayfa
Çizelge 4.23. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	75
Çizelge 4.24. Sticklefish verisi	77
Çizelge 4.25. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	77
Çizelge 4.26. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	79
Çizelge 4.27. Rahim serviksde karsinoma in situnun klasifikasyon verisi.....	82
Çizelge 4.28. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	82
Çizelge 4.29. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	83
Çizelge 4.30. Rahim serviksde karsinoma in situnun klasifikasyon verisi.....	86
Çizelge 4.31. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri	86
Çizelge 4.32. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri	87

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

κ	Cohen, Fleiss ve Conger'in kappa katsayısını
α	Krippendorff'un alf akatsayısı
π	Scott'un pi katsayısı
AC_2	Gwet'in ikinci sıra uyum katsayısı
n	Tüm denek sayısı
w_{kl}	Ağırlık değişkeni
T_w	Tüm ağırlık değerleri olan w_{kl} 'lerin toplamı
p_a	Gözlenen uyum
p_e	Beklenen uyum
f	Örnekleme kesri
$\bar{w}_{gk}^{(i)}$	g değerlendiricisi tarafından x_l skoru atanmış
$\hat{\alpha}_K$	Krippendorff'un α katsayısının tahmin edicisi
p_a^*	Krippendorff'un α katsayısının gözlenen uyumu
$\hat{\kappa}'_G$	Gwet AC_2 katsayısının tahmin edicisi
k, l	Değerlendiriciler tarafından atama yapılan kategoriler
q	Kategori sayısı
$\hat{\kappa}_C$	Cohen ve Conger kappa katsayısının tahmin edicisi
$\hat{\kappa}_{BP}$	Brennan-Prediger B-P katsayısının tahmin edicisi
$v(\hat{\kappa}_{BP})$	Brennan-Prediger B-P katsayısının varyans kestiricisi
$v(p_a)$	Yüzde uyumun varyansı
$\bar{u}$	u_i değerlerinin ortalaması
$v(\hat{\kappa}_C)$	Cohen ve Conger kappa katsayısının varyans kestiricisi
$v(\hat{\kappa}_S)$	Scott'ın π katsayısının varyans kestiricisi
$v(\hat{\kappa}_q)$	Brennan-Prediger katsayısının varyans kestiricisi
$\hat{\kappa}_S$	Scott'ın π katsayısının tahmin edicisi
π_k^*	Nispi oranların ortalamasının ortalaması

Simgeler**Açıklamalar**

$b_k^{(i)}$	$b_{+k}^{(i)}$ ve $b_{k+}^{(i)}$ değerlerinin ortalaması
r_i	i. deneği skorlayan değerlendirici sayısı
r	Değerlendirici sayısı
$\bar{r}$	r_i 'lerin ortalama değeridir. Bir deneği skorlayan değerlendiricilerin ortalama sayısıdır.
p'_e	Ağırlıklandırılmış gözlenen uyum nispi oranı
p_{k+}	A değerlendiricisinin x_k skorunu atadığı deneklerin oranı
n_{k+}	A değerlendiricisinin x_k skorunu atadığı deneklerin sayısı
p_A	A değerlendiricisi tarafından skorlanan deneklerin oranı
n_A	A değerlendiricisi tarafından skorlanan deneklerin sayısı
p_{+l}	B değerlendiricisinin x_l skorunu atadığı deneklerin oranı
n_{+l}	B değerlendiricisinin x_l skorunu atadığı deneklerin sayısı
p_B	B değerlendiricisi tarafından skorlanan deneklerin oranı
n_B	B değerlendiricisi tarafından skorlanan deneklerin sayısı
θ	En az bir değerlendirici tarafından skorlanan deneklerin yüzdesi
p_{AB}	Her iki A ve B değerlendiricisi tarafından skorlanan deneklerin oranı
n_{AB}	Her iki A ve B değerlendiricisi tarafından skorlanan deneklerin sayısı
$\bar{p}_{+k}$	B değerlendiricisinin atadığı x_k skorunu ağırlıklandırılmış toplamı
$\bar{p}_{l+}$	A değerlendiricisinin atadığı x_l skorunu ağırlıklandırılmış toplamı

Simgeler	Açıklamalar
θ_A	A değerlendiricisinin skorlamış olduğu deneklerin yüzdesi
θ_B	B değerlendiricisinin skorlamış olduğu deneklerin yüzdesi
π'_k	k. kategoride A ve B değerlendiricilerinin nispi oranlarının ortalaması
$\bar{\pi}_k$	k. kategoride ağırlıklandırılmış nispi oranların ortalaması
r_{il}	i. deneğe belli bir x_l skorunu atayan değerlendiricilerin sayısı
r_{ik}	i. deneğe belli bir x_k skorunu atayan değerlendiricilerin sayısı
r_{ik}^*	Tüm q kategoriler boyunca r_{il} 'lerin ağırlıklandırılmış toplamı
n_{gk}	g değerlendiricisi ve x_k skoruyla ilişkilendirilen deneklerin sayısı
n_g	g değerlendiricisi tarafından skorlanan deneklerin sayısı
n_{gl}	g değerlendiricisi ve x_l skoruyla ilişkilendirilen deneklerin sayısı
n'	Üç veya daha fazla değerlendirici tarafından skorlanan deneklerin sayısı
p'_{kl}	A değerlendiricisi tarafından k ve B değerlendiricisi tarafından l kategorisine atanan denek oranının, her iki değerlendirici tarafından skorlanan deneklerin oranına nispi oranı
p_{kl}	A ve B değerlendiricileri tarafından sırasıyla k ve l kategorilerine atanan deneklerin oranı
n_{kl}	A ve B gibi iki değerlendirici tarafından sırasıyla k ve l kategorilerine atanan denek sayısı
p'_{k+}	A değerlendiricisinin tüm atama yaptığı skorlar içerisinde x_k skorunu atadığı deneklerin nispi oranı

Simgeler**Açıklamalar** p'_{+l}

B değerlendiricisinin tüm atama yaptığı skorlar içerisinde x_l skorunu atadığı deneklerin nispi oranı

 p'_a

A ve B değerlendiricilerinin sırasıyla atadığı x_k ve x_l skorlarının ağırlıklandırılmış toplamı

 ε_i

Her iki değerlendirici de i. deneği sınıflamışsa 1, değilse 0 değerini alan değişken

 $\delta_{kl}^{(i)}$

A ve B değerlendiricileri i. deneği k ve l kategorilerine atadıysa 1, değilse 0 değerini alan ikili değişken

 $\varepsilon_{i.}$

A değerlendiricisi i. deneği skorlamışsa 1, değilse 0 değerini alan ikili değişken

 $\delta_{k+}^{(i)}$

A değerlendiricisi i. deneği k kategorisine sınıfladıysa 1, değilse 0 değerini alan ikili değişken

 $\varepsilon_{.i}$

B değerlendiricisi i. deneği skorlamışsa 1, değilse 0 değerini alan ikili değişken

 $\delta_{+k}^{(i)}$

B değerlendiricisi i. deneği k kategorisine sınıfladıysa 1, değilse 0 değerini alan ikili değişken

 $\hat{\kappa}_{(g,h)}$

g ve h olarak etiketlenen 2 değerlendirici arasındaki şansa bağlı olmayan (şans düzeltmeli) uyum katsayısı

 $\hat{\kappa}'_q$

Çoklu değerlendirici durumundaki Brennan-Prediger katsayısının tahmin edicisi

 $\delta_{gl}^{(i)}$

g değerlendiricisi i. deneği l kategorisine sınıfladıysa 1, değilse 0 değerini alan ikili değişken

Kısaltmalar**Açıklamalar****Std Hata**

Standart hata

G.A.

Güven aralığı

 κ

Cohen ve Conger kappa katsayılarının sembolü

 π

Scott'ın pi katsayısının sembolü

 α

Krippendorff'un alfa katsayısının sembolü

Kısaltmalar**Açıklamalar****B-P**

Brennan-Prediger katsayısının sembolü

Kripp. α Krippendorff'un α katsayısı

1. GİRİŞ

Yazıdan ve hatta antik çağlardan önce ortaya çıktığı bilinen bilim, evrenin yapısını deney veya gözlemlerle sistematik bir şekilde incelemeyi kapsayan pratik çalışmaları kapsamaktadır. Bilimsel yaklaşıma göre var olan her şey ölçülebilirdir. Deney veya gözlem yapılmadan elde edilen bir bilginin bilimsel olduğunu söyleyebilmek oldukça güçtür. Bilimin varlığına ve yöntemine olan şiddetli ihtiyaç, tarih öncesinden günümüze kadar gelen çalışmalara neden olmuştur. Bilimsel çalışmalar kuşkusuz kanıta dayalı ve tutarlı sonuçlar vermelidir. Bilimsel bakış, herhangi bir kanıta dayanmayan ve tutarsız bilgileri kabul etmemeyi gerektirir. Bilgiler, çeşitli ölçüm, deney veya gözlemlerden elde edilebileceği için ölçümün güvenilirliğini sağlamak gerekir.

Literatürde, güvenilirlik özellikle ölçüm güvenilirliği alanında öne çıkan kavramlardan biridir. Bademci (2011) güvenilirliği, belirli bir evrene veya örnekleme uygulanmış bir test ya da ölçme aracından elde edilmiş ölçümlerin tutarlılığı veya tekrarlanabilirliği şeklinde tanımlamaktadır. Bu bağlamda, bir testin ya da bir ölçme aracının yaptığı ölçümlerin tutarlılığını veya tekrarlanabilirliğini değerlendirmek için çeşitli yöntemlerin gelişimine ihtiyaç duyulmuştur.

Ölçüm için kullanılan testlerin, ölçeklerin vb. ölçüm güvenilirliğini test etmek için çeşitli yöntemler geliştirilmiştir. Cronbach alfa, Kuder-Richardson 20-21, test-retest, korelasyon katsayıları, sürekli veriler için sınıf içi korelasyon katsayısı ICC, kategorik veriler için uyum ve ilişki katsayıları vb. ölçüm güvenilirliğini hesaplamak için literatürde sıklıkla kullanılanlardan bazılarıdır. Bu ölçümler yapılırken örneklerden elde edilen parametre tahminlerinin şüphesiz ki istatistiksel olarak yığın parametresinin iyi bir temsili olma özelliği aslında ölçümlerden elde edilen sonuçları güvenilir yapan şeydir.

Kılıç (2009) Ölçümlerin Uyumluluğu ve Tıptaki Uygulamaları adlı yüksek lisans tezinde, ilişki ile uyum arasındaki farkı, “eğer iki değerlendirici arasında uyumluluk varsa aynı zamanda ilişkinin de var olduğu söylenebilir, fakat ilişkinin mükemmel olması uyumluluğunda mükemmel olduğu anlamına gelmez” şeklinde açıklanmıştır. Bu bağlamda, bu tezde değerlendiricilerin verdiği kararların uyumunu ölçmek esas alındığı için ilişki katsayılarına değinilmemiş ve sadece uyumu ölçen katsayılar dikkate alınmıştır.

Sınıflama ölçek düzeyinde ve iki değerlendiricili durum için oldukça çok sayıda çalışma bulunmasına rağmen, sıralı, aralıklı ve oransal ölçek düzeyleri ve çoklu değerlendiricili durumlar için geliştirilen yaklaşımlar nispeten sınırlı sayıdadır. Literatürde bahsedilen en eski şansa bağlı olmayan uyum katsayısı Benini tarafından 1901 yılında bahsedilmiştir. Şansa bağlı uyum meselesini çözmek için diğer erken dönem çalışmaları Guttman (1945), Bennett, Albert ve Goldstein (1954), Holley ve Guilford (1964), Maxwell (1977), Janson ve Vegelius (1979) ve Brennan-Prediger (1981) gibi birbirinden bağımsız olarak aynı katsayıyı geliştiren fakat farklı isim veren bilim adamlarına aittir. Bu basit katsayıya, literatürde sıklıkla Brennan-Prediger katsayısı olarak atıfta bulunulur ve

$$(p_a - 1/q) / (1 - 1/q) \quad (1.1)$$

oranıyla verilir, burada p_a uyum yüzdesi ve q , skarlama ölçeğindeki nominal kategorilerin sayısıdır. Brennan-Prediger (1981) gibi birçok yazar B-P katsayısını iki değerlendiricili ve iki kategorili durum için önermeden önce, kendi katsayılarını iki değerlendiricili ve q tane rasgele seçilmiş kategorili durum için önermişlerdir. Üç veya daha fazla değerlendiricinin olduğu durumlar için de bu katsayı genişletilebilirdi. Holley ve Guildford (1964), her ne kadar diğer yazarların çeşitli makalelerinde önceden bahsedilse de, değerlendiriciler arası güvenilirliği hesaplamanın bir yolu olarak bu katsayıyı resmi olarak çalışan ilk kişilerdir. Onlar bu katsayıya *G-index* adını vermişlerdir. Geliştirdikleri *G-index* katsayısı literatürde birçok yazar tarafından, uyum katsayısının büyüklüğünü arttırmak için araştırmacıların bu katsayıyı hesaplarken, kukla kategoriler eklenebileceği endişesi ile eleştirilmiştir.

1940'ların sonu ve 1950'lerin başlangıcından beri literatürde birtakım değerlendiriciler-arası güvenilirlik katsayıları önerilirken, Cohen(1960), şüphesiz ki en yaygın kullanılacak olan Kappa katsayısını önermiştir. Cohen, ilk olarak, nominal ölçekli veriler için 2 değerlendirici ve 2 kategoride kullanılacak şekilde bir katsayı sunmuştur. Literatürde ve pratikte kullanımı oldukça yaygın olan bu uyum katsayısı, yine literatürde oldukça sık bahsedilen paradoksları nedeniyle araştırmacıların dikkatini çekmiştir. Literatürde “Kappa Paradoksu” olarak nitelendirilen bu dezavantaj, birçok yazar tarafından irdelenmiş ve alternatif katsayıların geliştirilmesine neden olmuştur(Feinstein ve Cicchetti, 1990, Gwet, 2008a).

Değerlendiriciler arası güvenilirlik literatüründe, erken dönem çalışmaları sınıflama ölçekli skorlamalar için çeşitli katsayılar içerirken, sıralı ölçekli skorlamalara geldiğinde, bu katsayıların kullanılmasının hatalı ölçümlere sebep olduğu tespit edilmiştir. Bu nedenle, mevcut katsayılar sıralı, aralıklı ve oranlamalı ölçek tipindeki skorlamalar için genelleştirilmiştir. İlk genelleştirme, Cohen (1968) tarafından ordinal ölçekli veriler içinde kullanılabilecek “Ağırlıklandırılmış Kappa Katsayısı” ile sunulmuştur. Cohen’in kullandığı ağırlık kavramı, sıralı ölçekleri, belli kategorilere ağırlıklar atayarak nominal ölçeğe çeviren bir yöntemdir. Ardından, bu genelleştirme fikri, sıralı, aralıklı ve oranlamalı ölçek tiplerini de kapsayacak şekilde, ilk olarak Berry ve Mielke (1988) tarafından ortaya atılmış olup, daha sonra Janson ve Olsson (2001, 2004) tarafından kayıp verileri de ele alacak biçimde şekillendirilmiştir.

Berry ve Mielke (1988) ve Janson ve Olsson (2001), Cohen (1968) tarafından önerilmiş ağırlıklandırılmış kappa katsayısını sıralı ve aralıklı veriye indirgeyen kuadratik ağırlıklar kullanmışlardır. Bu ağırlıkların kullanılması benzer şekilde Scott’ın π ’sini (Scott, 1955), Brennan-Prediger istatistiğini (Brennan ve Prediger, 1981), Krippendorff’un α katsayısını (Krippendorff, 1970, 1978, 2004) ve Gwet’in AC_1 ’inin (Gwet, 2001) de sıralı, aralıklı ve oranlamalı skorlar için kullanılabilmesine olanak tanımıştır. Ayrıca, Conger (1980) de kappa istatistiğini 3 veya daha fazla değerlendiricili duruma genişletmiştir. Bunu Janson ve Olsson (2001) yaklaşımını izleyerek gerçekleştirmiştir.

Bu geniş uygulama alanında, bu tezin konusu olarak sıralı ölçeklerde bir fenomene karar veren iki ya da daha fazla değerlendiricinin arasındaki karar verme uyumluluğunun güvenilirliğini inceleme seçildi. Bu konunun bu tezde ele alınması öncelikli olarak son 50 yıldır tıp, eğitim, psikometri vb. alanlarda belli değerlendiricilerin uyumu ile değerlendirilen fenomenlerin çoğalmasına ve bu alanlarda sağlam istatistiklere olan ihtiyaca dayanmaktadır.

Bu ihtiyaç, değerlendiricilerin denekler üzerinde verdikleri kararların birbirleri arasındaki uyumunun ne kadar güvenilir olduğunun tespitidir. Belli bir hastalıkla ilgili, herhangi bir ilde yaşayan nüfusun, hastalığa karşı eğilimini teşhis edecek olan doktorlar ele alınsın. Böyle bir deneyde, deneyin yapıldığı ildeki doktorların tamamının içerisinde r adet doktorun yaptığı teşhisin, tüm doktorlar bu deneye katılsaydı meydana gelebilecek teşhisin iyi bir temsili olup olmadığı önemli bir durumdur. Bunun sebebi, R adet doktorun içerisinde tesadüfen seçilen r adet doktorun kararlarının, eğer farklı bir r adet doktor seçilseydi

verilecek kararlardan ne kadar farklı olacađının bilinmek istenmesidir. Deđerlendirici temelli bu varyasyon sebebiyle kestirilen uyum katsayısının temsil ettiđi uyumun gúcüne ancak uyum katsayısının gerçekteki uyumu iyi bir şekilde temsil etmesiyle gúvenilebilir. Deđerlendiriciler arası gúvenirlilik çalıřmasına önderlik eden iřte bu problemin çözölme ihtiyacıdır. Bu tezin amacı da, literatürde bu problemin çözölüne iliřkin nominal ölçekte geliřtirilen yöntemlerin sıralı ölçek düzeyindeki skorlamalar için geniřletilmiş hallerini ele almak ve karşılařtırmaktır.

2. SIRALI ÖLÇEK SKORLARI İÇİN DEĞERLENDİRİCİLER-ARASI GÜVENİLİRLİK ÖLÇÜMÜ YÖNTEMLERİ

Değerlendiriciler-arasındaki uyum değerlendirilirken ilk olarak dikkate alınan özellik değerlendirici sayısıdır. Uyum katsayıları iki ve ikiden fazla değerlendirici olduğu durumlarda farklı hesaplanmaktadır. Literatürde, nominal ölçekler için değerlendiriciler arası uyumu hesaplayan çok sayıda katsayı olmasına rağmen, sıralı ölçeklerde değerlendirme yapan 2 veya daha fazla değerlendirici için sınırlı sayıda uyum katsayısı mevcuttur. En çok kullanılan uyum katsayıları; Cohen'in ağırlıklandırılmış kappası (Cohen, 1968), Krippendorff'un α katsayısı (Krippendorff, 1970, 1978, 2004), Fleiss'in kappası (Fleiss, 1971) ve Conger'in kappası (Conger, 1980) dir. Son yıllarda bu katsayılara alternatif olarak Gwet (2001) tarafından AC_1 katsayısı geliştirilmiş olup ve yine literatürdeki diğer katsayılardan bazıları Jonson ve Olsson (2001) tarafından geliştirilen uzaklık fonksiyonu kullanarak eksik skorlamaları da kapsayacak şekilde Gwet (2014) tarafından genelleştirmiştir.

2.1. Sıralı Ölçek Kavramı

Sıralı ölçekler, ilgilenilen özellik bakımından, birimlerin birbirinden farklı özelliklere sahip olmalarının yanı sıra, ilgili özellik bakımından birimlerin hangi büyüklük sırasına ait olduğunu belirlemek amacıyla kullanılırlar. Kanser tedavisini planlamak için, hekimin hastalığın evresini bilmesi gerekir. Hastalığın evresi, tümörün boyutu ve ne kadar yayıldığıyla ilgilidir. Bu amaçla patolojik hastalığı Evre I, Evre II, Evre III ve ya Evre IV şeklinde kategorilendirir. Kategoriler düşükten yükseğe doğru sıralandırıldığı için bu tipteki bir ölçeğe sıralı ölçek adı verilir.

Sıralı ölçeklerde kategoriler küçükten büyüğe doğru sıralanabildiği gibi numaralamanın yapılmadığı ancak belli bir büyüklük sırasına sahip kategorilerin olduğu durumlarda vardır. Örneğin, insanlar “Beden kitle indeksine” göre “Zayıf”, “Normal”, “Hafif kilolu” ve “Obez” olarak da kategorilendirilebilir. Bu örnekteki kategoriler de küçükten büyüğe doğru bir sıralama belirttiği için sıralı ölçektir. Sıralı ölçekteki temel nokta, kategorilerin bir büyüklük sırasına göre sıralanmış olmasıdır.

2.2. Uyum Kavramı

Değerlendiriciler arası güvenilirliği değerlendirmede dikkate alınan uyum kavramının literatürde yaygın olarak kabul edilmiş bir tanımı yoktur. Ancak en basit tanımlamayla “kategorizasyonların kesişiminin büyüklüğü” şeklinde ifade edilmektedir (Agresti, 1988; Gwet, 2008a). Bu kesişim (uyum) ne kadar büyükse güvenilirliğin o kadar büyük çıkması beklenir. Bu tanım genellikle nominal ölçekler için yapılmıştır. Sıralı ölçeklerde ise uyumsuzluk da kısmi uyum olarak dikkate alındığı için tanımın genişletilmesine ihtiyaç duyulmuştur.

Deneklerin önceden tanımlanmış kategorilere/sınıflara atanması birçok alanda (sağlık, finans, eğitim, psikoloji, yazılım, dilbilim gibi) oldukça yaygın olarak başvurulmuş bir yöntemdir. Örneğin hastalar hastalık durumlarına göre, kredi başvuruları risk değerlendirmesine göre, kişiler depresyon durumlarına göre değerlendiriciler tarafından kategorilere atanabilir. Yapılan bu sınıflama sürecinin güvenilir olması için iki veya daha fazla değerlendiricinin aynı denek setini değerlendirmesine ihtiyaç duyulur. Bu durumda güvenilirliğin değerlendirilebilmesi için değerlendiricilerin arasındaki uyumun ölçülmesi gerekir.

Eğer uyum yüksekse, o zaman araştırmacı değerlendirici faktörü tarafından etkilenmiş bir kategorizasyondan endişelenmeksizin aynı konudaki farklı bir çalışmada her iki değerlendiriciyi değişimli olarak kullanabilir. Örneği 2 değerlendirici kendi arasında yüksek bir uyum göstermiş ise, araştırmacı, yeni bir değerlendirmede, benzer değerlendirmeler yapacakları beklendiği için ikisinden birini seçebilir. Buna değerlendiricilerin değiştirilebilirliği denmektedir ve bu kavram değerlendiriciler arası güvenilirliğin önemini doğrulayan şeydir. Eğer değiştirilebilirlik garanti edilirse, o zaman deneklerin sınıflandırıldıkları kategoriler değerlendiricinin ne ürettiği sorgulanmadan güvenle kullanılabilir.

Uyum, literatürde çeşitli uyum katsayıları ile ölçülmektedir ve değerlendiriciler arası güvenilirlik çalışmasına katılan değerlendirici, kategori ve denek sayısından etkilenmektedir.

Uyum katsayısı formüllerinde, uyumun gözlenen oranı p_a “yüzde uyum” ve uyumun beklenen oranı p_e “yüzde şans uyumu” ile simgelenmektedir.

2.2.1. Şansa bağlı uyum kavramı(p_e)

Değerlendiricilerin verdiği kararların uyumunun incelendiği bir değerlendiriciler arası güvenilirlik çalışmasında, uyumun kaynağının tamamen şans olma ihtimali mevcuttur. Söz gelimi, iki değerlendirici ve iki kategoriden oluşan bir deneyde, her bir değerlendirici belli bir deneği 0,5 olasılıkla iki kategoriden birine atamalıdır. Bu, iki değerlendiriciden birinin, deneğin hangi kategoriye ait olduğunu bilmediği ve deneği risk olarak mümkün tesadüfi iki kategoriden birine 0,5 gibi (50;50 oranı) eşit bir olasılıkla sınıflandırdığı anlamına gelebilmektedir. Şansa bağlı uyum kavramı, bu bağlamda tesadüfi skorlamalara bağlıdır.

Sınıflama sürecinde değerlendiriciler, gözlenen marjinal olasılıklara göre bir sınıflama yaptıklarında şans yoluyla uyuştukları söylenebilir. Cohen(1960), nominal ölçekler için kappa uyum katsayısını geliştirdiğinde, yukarıda bahsedilen şansa bağlı uyum kavramını ölçebilmek için marjinal sınıflama olasılıklarına dayanan p_e gibi bir yüzde şans uyumu geliştirmiştir. Skorlamalar tamamen tesadüfi olduğunda p_a yüzde uyumu ile p_e yüzde şans uyumunun birbirlerine yakın olması beklenir. Bu, matematiksel olarak, uyum katsayısının 0'a oldukça yakın bir değer alması ve uyumun olmaması anlamına gelir. Buradan teorik olarak denebilir ki, değerlendiriciler arası bir güvenilirlik çalışmasında uyum ölçülürken, değerlendiricilerin kararlarında ortaya çıkan tesadüfi bir uyum, aslında uyumun olmadığına bir göstergesi olacaktır.

Böyle şansa bağlı bir uyum, uyum hesaplamalarında istenmemektedir. Bu nedenle, literatürdeki çeşitli katsayıların hepsi farklı p_e formülasyonlarına sahiptir ve bu formülasyonlar, tesadüfen meydana gelebilecek bir uyumu, uyum hesaplamasına minimum düzeyde dâhil etmeyi ya da hiç dâhil etmemeyi amaçlamaktadır.

2.2.2. Kısmi uyum kavramı

Değerlendiriciler arasındaki değerlendirme uyumsuzlukları “kısmi uyum” olarak tanımlanmaktadır (Cohen, 1968; Landis ve Koch, 1977a; Gwet 2014:74).

Örneğin, bir diyetisyenin hem hastalardan kendilerini hem de kendisinin hastaları “Zayıf”, “Normal”, “Aşırı kilolu” ve “Obez” olarak isimlendirilen kategorilerden birine atadığı bir deney düşünülün. Hastanın kendisini “Aşırı kilolu” olarak ve diyetisyenin bu hastayı

“Obez” olarak değerlendirdiği varsayılırsa, bu iki sınıflama arasında bir uyumsuzluk olduğu görülür. Başka bir hasta kendisini “zayıf” olarak ve diyetisyen bu hastayı “Obez” olarak değerlendirdiğinde yine bu iki sınıflama arasında da bir uyumsuzluk vardır. Burada ikinci durumdaki (birbirinden uzak olan sıralı iki kategori) uyumsuzluğun ilk duruma (ardışık iki sıralı kategori) göre daha büyük bir uyuşmazlığa sahip olduğuna dikkat edilmelidir. Değerlendiriciler arası güvenilirlik çalışmalarında bu tür uyuşmazlıklara “kısmi uyum” denir ve aynı ölçüde değerlendirilmemelidir. Kısmi uyumun değerlendirmesi bu uyuşmazlıklara atanacak çeşitli ağırlıklar ile yapılır.

2.3. Ağırlıklandırma Kavramı

Sıralı ölçeklerde hesaplanacak uyum katsayısındaki uyumlar kadar kısmi uyumların da değerlendirilmesi için ağırlıklandırma yöntemi kullanılmaktadır.

Ağırlıklandırma yöntemi, uyum katsayısında etki etmesi istenen kısmi uyumların hangisinin dikkate alınacağını belirlenmesi ve hesaplanması sürecidir. Bazı “küçük” uyumsuzluklar, değerlendiricilerin aynı denek hakkındaki değerlendirmelerinde hemen hemen benzer bir yargıya sahip olduklarını gösterebilir. Bu nedenle mükemmel uyumlara 1, kısmi uyumlara ise 1’den daha küçük ağırlık değerleri atanabilir. Eğer bazı kısmi uyumlar mükemmel uyumlar kadar önemli ise, bu kısmi uyumlara da örneğin 1’e yakın 0,9 gibi bir ağırlık atanabilir.

Uyum katsayılarının sıralı, aralıklı ve oranlamalı skorlar için genişletilmesini sağlayan çeşitli ağırlıklandırma yöntemleri vardır. Bunlar, identity (özdeş), radikal, doğrusal, kuadratik, basit sıralı, oransal, dairesel, bipolar ağırlıklar olarak listelenebilir. Ancak, en iyi ağırlık yönteminin seçimi belli bir kurala dayanmamaktadır. Sıralı ölçekler için bu ağırlık yöntemlerinden “Kuadratik” ve “Basit sıralı” ağırlık yöntemi yaygın olarak kullanılmaktadır.

2.3.1. Kuadratik ağırlıklar

Cohen (1968) ağırlıklı kappa katsayısı olarak bilinen uyum katsayısını önerdiğinde, bu önerdiği katsayısı aslında karesi alınmış öklid uzaklıklarının bir fonksiyonuydu. Karesi alınmış Öklid uzaklıkları, kuadratik ağırlıkların oluşturulmasında temel teşkil etmiştir. A

değerlendiricisinin x_k ve B değerlendiricilerin x_l skorunu atadığı herhangi bir denek için atama yapılan iki kategorinin kendileriyle ilişkilendirilen bir w_{kl} ağırlığı vardır ve aşağıdaki gibi tanımlanır:

$$w_{kl} = \begin{cases} 1 - \frac{(x_k - x_l)^2}{(x_{max} - x_{min})^2}, & k \neq l \\ 1, & k = l \end{cases} \quad (2.1)$$

burada x_{max} ve x_{min} sırasıyla en büyük ve en küçük skorlardır. Eş. 2.1 tarafından tanımlanan ağırlık kümesi literatürde “Kuadratik Ağırlıklar” olarak bilinir.

Sıralı skorlamada kategoriler artan sırada rakamlarla ifade edilebildiği gibi yine bir sıralama bildirecek şekilde “düşük”, “orta” ve “yüksek” gibi alfabetik olarak da ifade edilebilir. Alfabetik olan kategorilere de, ağırlık hesaplanırken artan sıralamasına göre sıralı olarak 1, 2 ve 3 gibi tamsayı değerleri atanır.

2.3.2. Basit sıralı ağırlıklar

Basit sıralı ağırlıkların hesaplanmasında aşağıdaki formül kullanılır:

$$w_{kl} = \begin{cases} 1 - M_{kl}/M_{1q}, & k \neq l \\ 1, & k = l \end{cases} \quad (2.2)$$

$M_{kl} = \#\{(i,j): \min(k,l) \leq i < j \leq \max(k,l)\}$, $\min(k,l)$ ve $\max(k,l)$ arasında oluşturulabilecek $i < j$ olmak üzere (i,j) çiftlerinin sayısını temsil etmektedir. Burada kategoriler sırasıyla en düşük (1) en yüksek (q) olarak varsayılmaktadır. M_{kl} aşağıdaki gibi hesaplanır:

$$M_{kl} = \binom{\max(k,l) - \min(k,l) + 1}{2} \quad (2.3)$$

M_{kl} , $\max(k,l) - \min(k,l) + 1$ in 2’li kombinasyonudur.

Yukarıdaki eşitliklerden elde edilecek basit sıralı ağırlıklar kümesi, kuadratik ağırlıklar hesaplanırken kullanılan öklid uzaklık kavramına dayanmayıp, skorlamaların sıra temelini kullanır.

Ağırlıkların hesaplanış şekilleri EK-1’de gösterilmiştir.

2.4. İki Değerlendiricili Durum için Uyum Katsayıları

İki değerlendirici için değerlendiriciler-arası uyumu gösteren çeşitli uyum katsayıları aşağıda listelenmiştir.

Cohen’in ağırlıklandırılmış kappa katsayısı (1968)

Scott’ın ağırlıklandırılmış pi katsayısı (1955)

Ağırlıklandırılmış Brennan-Prediger B-P katsayısı (1981)

Krippendorff’un α katsayısı (1970;1978; 2004)

Gwet’in AC₂ katsayısı (2001)

Bu uyum katsayılarının koşullu çıkarım ve koşulsuz çıkarım durumunda sahip oldukları varyans kestiricileri mevcuttur. Deneklerin rasgele seçilip, değerlendiricilerin rasgele seçilmediği veya ilgilenilen tüm değerlendiricilerin bir değerlendiriciler-arası güvenilirlik çalışmasına dahil edildiği durumlarda koşullu varyans formülleri kullanılmaktadır. Bu çalışmada, koşullu varyans kestiricilerinin formülleri de verilmiştir.

Değerlendiriciler arası güvenilirlik çalışmasında bazı deneklerin her iki değerlendirici tarafından skorlanmasının aksine tek bir değerlendirici tarafından skorlanması durumunda veri, kayıp skorlamalar içeriyor demektir. Bu çalışmada, ilgili koşullu varyans formülasyonunun yanı sıra, kayıp veri durumunda kullanılabilecek varyans kestirimi de verilmiştir.

Bu formülasyonların hepsinde temel bir kappa katsayısı formülü bulunmaktadır. Formül Eş. 2.4’teki gibidir:

$$\kappa = \frac{p_a - p_e}{1 - p_e} = \frac{\text{Gözlenen uyum} - \text{Şansa bağlı uyum}}{1 - \text{Şansa bağlı uyum}} \quad (2.4)$$

2.4.1. Cohen’in ağırlıklandırılmış kappa katsayısı

Cohen (1968) kappa katsayısı karesi alınmış öklid uzaklıklarının bir fonksiyonudur. Sıralı ölçek ve hiçbir eksik skorlaması olmayan tam bir veri seti ile ilgilenildiğinde, Cohen (1968)

kappa katsayısı kullanılabilir. Ancak, eksik skorlama olması durumunda, Cohen (1968) tarafından geliştirilen ağırlıklandırılmış kappa katsayısı kullanılamamaktadır. Örneğin, A ve B değerlendiricilerin sırasıyla k ve l kategorilerine sınıflandırdığı deneklerin oranı, her iki değerlendiricinin skorlamış olduğu deneklerle ilgili olarak değerlendirilmelidir. Eğer bir denek, 2 değerlendiriciden yalnızca biri tarafından skorlandıysa, o zaman bu oran kappa hesaplanmasından hariç tutulmuştur. Bu denekleri hariç tutmak, Kappa katsayısının olduğundan az hesaplanmasına neden olmuştur. Bu nedenle, Gwet (2014:80) Kappa'nın daha genel bir formülasyonunu önermiştir. Kappa'nın öklid uzaklığının bir fonksiyonu olduğuna dair detaylı açıklama EK-2'de verilmiştir.

Değerlendirici sayısı 2 ile kısıtlı olduğunda sıralı veriler için Kappa katsayısı aşağıdaki genel forma sahiptir:

$$\hat{\kappa}_C = \frac{p_a - p_e}{1 - p_e}, \quad p_a = \sum_{k,l}^q w_{kl} p'_{kl} \quad \text{ve}, \quad p_e = \sum_{k,l}^q w_{kl} p'_{k+} p'_{+l} \quad (2.5)$$

Eş. 2.5, literatürde Cohen (1968) tarafından önerilen ağırlıklı Kappa katsayısı olarak tanımlamaktadır. Cohen'in de (1968) ağırlıklı Kappa'yı tanımlarken önerdiği gibi, araştırmacılar, deney tasarımını en iyi tarif edebilecek ağırlıkların özel bir kümesini iyi şekilde belirleyebilirler.

Burada;

$$p'_{kl} = p_{kl} / p_{AB} \quad (2.6)$$

$$p_{kl} = n_{kl} / n \quad (2.7)$$

$$p_{AB} = n_{AB} / n \quad (2.8)$$

$$p'_{k+} = p_{k+} / p_A \quad (2.9)$$

$$p_{k+} = n_{k+} / n \quad (2.10)$$

$$p_A = n_A / n \quad (2.11)$$

$$p'_{+l} = p_{+l}/p_B \quad (2.12)$$

$$p_{+l} = n_{+l}/n \quad (2.13)$$

$$p_B = n_B/n \quad (2.14)$$

κ varyans kestiricisi

$$v(\hat{\kappa}_C) = \frac{1-f}{n(1-p_e)^2} \{ \sum_{k,l}^q p_{kl} [w_{kl} - (1 - \hat{\kappa}_C)(\bar{p}_{+k} + \bar{p}_{l+})]^2 - [p'_a - 2(1 - \hat{\kappa}_C)p'_e]^2 \} \quad (2.15)$$

Burada;

$$f: f = n/N \text{ 'i örneklem kesri}^1 \text{ olarak tanımlamıştır.} \quad (2.16)$$

p_e : Eş. 2.5'te tanımlandığı gibidir.

p_{kl} : Eş. 2.7'de tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$\hat{\kappa}_C$: Eş. 2.5'te tanımlandığı gibidir.

$$\bar{p}_{+k} = \sum_{l=1}^q w_{kl} p_{+l} \quad (2.17)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p_{+l} : Eş. 2.13'te tanımlandığı gibidir.

$$\bar{p}_{l+} = \sum_{k=1}^q w_{kl} p_{k+} \quad (2.18)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p_{k+} : Eş. 2.10'de tanımlandığı gibidir.

$$p'_a = \sum_{k,l}^q w_{kl} p_{kl} \quad (2.19)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p_{kl} : Eş. 2.7'de tanımlandığı gibidir.

$$p'_e = \sum_{k,l}^q w_{kl} p_{k+} p_{+l} \quad (2.20)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

¹Birçok çalışmada, denek popülasyon hacmi N bilinmemektedir. Böyle durumlarda, $f=0$ alınabilir. Bu, tüm deneysel amaçlar için örneklem kesrini önemsiz varsaymak anlamına gelmektedir.

p_{k+} : Eş. 2.10'da tanımlandığı gibidir.

p_{+l} : Eş. 2.13'te tanımlandığı gibidir.

Kayıp verili durum

$$v(\hat{\kappa}_C) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2 \quad (2.21)$$

Burada;

f : Eş. 2.16'da tanımlandığı gibidir.

$$u_i = u_{1i} + u_{2i} \quad (2.22)$$

$$u_{1i} = \frac{\varepsilon_i/\theta}{1-p_e} \sum_{k,l}^q w_{kl} (\delta_{kl}^{(i)} - p'_{kl}) \quad (2.23)$$

$$\varepsilon_i = \begin{cases} 1, & \text{eğer her iki değerlendirici de } i. \text{deneği sınıflamışsa} \\ 0, & \text{dd} \end{cases} \quad (2.24)$$

p_e : Eş. 2.5'te tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\delta_{kl}^{(i)} = \begin{cases} 1, & \text{eğer } A \text{ ve } B \text{ değerlendiricileri } i. \text{deneği} \\ & \text{sırasıyla } k \text{ ve } l \text{ kategorilerine sınıfladıysa} \\ 0, & \text{dd} \end{cases} \quad (2.25)$$

$$p'_{kl} = p_{kl}/\theta \quad (2.26)$$

$$u_{2i} = \frac{1-\hat{\kappa}_C}{1-p_e} \left(\sum_{k=1}^q \bar{p}_{+k} b_{k+}^{(i)} + \sum_{k=1}^q \bar{p}_{l+} b_{+k}^{(i)} \right) \quad (2.27)$$

$\hat{\kappa}_C$: Eş. 2.5'te tanımlandığı gibidir.

p_e : Eş. 2.5'te tanımlandığı gibidir.

$\bar{p}_{+k}$: Eş. 2.17'de tanımlandığı gibidir.

$\bar{p}_{l+}$: Eş. 2.18'de tanımlandığı gibidir.

$$b_{k+}^{(i)} = -\left(\frac{\varepsilon_{i.}}{\theta_A}\right) \left(\delta_{k+}^{(i)} - p'_{k+}\right) \quad (2.28)$$

$$\varepsilon_{i.} = \begin{cases} 1, & \text{eğer } A \text{ değerlendiricisi } i. \text{ deneği skorlamışsa} \\ 0, & \text{dd} \end{cases} \quad (2.29)$$

θ_A : A değerlendiricisinin skorlamış olduğu deneklerin yüzdesidir.

$$\delta_{k+}^{(i)} = \begin{cases} 1, & \text{eğer } A \text{ değerlendiricisi } i. \text{ deneği } k \text{ kategorisine sınıfladıysa} \\ 0, & \text{dd} \end{cases} \quad (2.30)$$

$$p'_{k+} = p_{k+} / \theta_A \quad (2.31)$$

p_{k+} : Eş. 2.7'de tanımlandığı gibidir.

$$b_{+k}^{(i)} = -\left(\frac{\varepsilon_{.i}}{\theta_B}\right) \left(\delta_{+k}^{(i)} - p'_{+l}\right) \quad (2.32)$$

$$\varepsilon_{.i} = \begin{cases} 1, & \text{eğer } B \text{ değerlendiricisi } i. \text{ deneği skorlamışsa} \\ 0, & \text{dd} \end{cases} \quad (2.33)$$

$$\delta_{+k}^{(i)} = \begin{cases} 1, & \text{eğer } B \text{ değerlendiricisi } i. \text{ deneği } k \text{ kategorisine sınıfladıysa} \\ 0, & \text{dd} \end{cases}$$

(2.34)

$$p'_{+l} = p_{+l} / \theta_B \quad (2.35)$$

p_{+l} : Eş. 2.13'te tanımlandığı gibidir.

$$\bar{u} = \sum_{i=1}^n u_i / n \quad (2.36)$$

u_i : Eş. 2.22'de tanımlandığı gibidir.

2.4.2. Ağırlıklandırılmış Scott'un π katsayısı

$\hat{\kappa}_S$, ağırlıklandırılmış π katsayısı olsun. Bu katsayı aşağıdaki gibi tanımlanır:

$$\hat{\kappa}_S = \frac{p_a - p_e}{1 - p_e}, \quad p_a = \sum_{k,l}^q w_{kl} p'_{kl} \quad \text{ve} \quad p_e = \sum_{k,l}^q w_{kl} \pi'_k \pi'_l \quad (2.37)$$

Burada,

w_{kl} : Eş 2.1 veya Eş. 2.2’de tanımlandığı gibidir.

p'_{kl} : Eş. 2.3’de tanımlandığı gibidir.

$$\pi'_k = (p'_{k+} + p'_{+k})/2 \quad (2.38)$$

p'_{k+} : Eş 2.6’da tanımlandığı gibidir.

p'_{+k} : Eş 2.9’da tanımlandığı gibidir.

$$\pi'_l = \pi'_k \quad (2.39)$$

π varyans kestiricisi

$$v(\hat{\kappa}_S) = \frac{1-f}{n(1-p_e)^2} \{ \sum_{k,l}^q p_{kl} [w_{kl} - (1 - \hat{\kappa}_S)(\bar{\pi}_k + \bar{\pi}_l)]^2 - [p'_a - 2(1 - \hat{\kappa}_S)p'_e]^2 \} \quad (2.40)$$

Burada,

f : Eş 2.16’da tanımlandığı gibidir.

p_e : Eş. 2.37’de tanımlandığı gibidir.

p_{kl} : Eş. 2.7’de tanımlandığı gibidir.

w_{kl} : Eş 2.1 veya Eş. 2.2’de tanımlandığı gibidir.

$\hat{\kappa}_S$: Eş 2.37’de tanımlandığı gibidir.

$$\bar{\pi}_k = (\bar{p}_{+k} + \bar{p}_{k+})/2 \quad (2.41)$$

$$\bar{\pi}_l = \bar{\pi}_k \quad (2.42)$$

$\bar{p}_{+k}$: Eş. 2.17’de tanımlandığı gibidir.

$\bar{p}_{l+}$: Eş. 2.18’de tanımlandığı gibidir.

p'_a : Eş 2.19’da tanımlandığı gibidir.

$$p'_e = \sum_{k,l}^q w_{kl} \pi'_k \pi'_l \quad (2.43)$$

w_{kl} : Eş 2.1 veya Eş. 2.2’de tanımlandığı gibidir.

π'_k : Eş. 2.38’de tanımlandığı gibidir.

π'_l : Eş. 2.39’da tanımlandığı gibidir.

Kayıp verili durum

$$v(\hat{\kappa}_S) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2 \quad (2.44)$$

Burada,

f : Eş. 2.16'da tanımlandığı gibidir.

u_i : Eş. 2.22'de tanımlandığı gibidir.

u_{1i} : Eş. 2.23'te tanımlandığı gibidir.

$$u_{2i} = 2 \frac{1-\hat{\kappa}_S}{1-p_e} \sum_{k=1}^q \pi_k^* b_k^{(i)} \quad (2.45)$$

$\hat{\kappa}_S$: Eş. 2.37'de tanımlandığı gibidir.

p_e : Eş. 2.37'de tanımlandığı gibidir.

$$\pi_k^* = (\bar{\pi}_{k+} + \bar{\pi}_{+l})/2 \quad (2.46)$$

$$\bar{\pi}_{k+} = \sum_{l=1}^q w_{kl} \pi'_l \quad (2.47)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

π'_l : Eş. 2.39'da tanımlandığı gibidir.

$$\bar{\pi}_{+k} = \sum_{l=1}^q w_{kl} \pi'_k \quad (2.48)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

π'_k : Eş. 2.38'de tanımlandığı gibidir.

$$b_k^{(i)} = (b_{+k}^{(i)} + b_{k+}^{(i)})/2 \quad (2.49)$$

$b_{+k}^{(i)}$: Eş. 2.32'de tanımlandığı gibidir.

$b_{k+}^{(i)}$: Eş. 2.28'de tanımlandığı gibidir.

2.4.3. Ağırlıklandırılmış Brennan-Prediger BP katsayısı

$\hat{\kappa}_{BP}$ ağırlıklandırılmış BP katsayısı olsun². Bu katsayı aşağıdaki gibi tanımlanır:

² Brennan-Prediger(1981)'in yalnızca sınıflama skorlarını alabilen ağırlıklandırılmamış bir uyum katsayısını ve hiçbir kayıt skorlaması olmayan tam veri setlerini ele aldığına dikkat edilmelidir. Gwet (2014:82) burada, aynı zamanda kayıp skorlamaları da ele alabilecek sıralı veriye bir genişletmeyi önermiştir.

$$\hat{\kappa}_{BP} = \frac{p_a - p_e}{1 - p_e}, \quad p_a = \sum_{k,l}^q w_{kl} p'_{kl} \quad \text{ve} \quad p_e = \frac{1}{q^2} \sum_{k,l}^q w_{kl} \quad (2.50)$$

Burada,

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p'_{kl} : Eş. 2.6'da tanımlandığı gibidir.

BP katsayısı Brennan-Prediger (1981) tarafından kapsamlı bir şekilde ele alınmış olup, Kappa ve π ile ilişkilendirilen paradokslara dirençli oluşu ile bilinmektedir. Gwet (2008a) tarafından, BP'nin şansa bağlı yüzde uyumu p_e 'nin zaman zaman tamamen şans yoluyla uyuşan değerlendiriciler için eğilimi az da olsa abartabildiği gösterilmiştir.

B-P varyans kestiricisi

$$v(\hat{\kappa}_{BP}) = \frac{1-f}{n(1-p_e)} (\sum_{k,l}^q w_{kl}^2 p_{kl} - p_a'^2) \quad (2.51)$$

Burada,

f : Eş. 2.16'da tanımlandığı gibidir.

p_e : Eş. 2.50'de tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p_{kl} : Eş. 2.7'de tanımlandığı gibidir.

p_a' : Eş. 2.19'da tanımlandığı gibidir.

Kayıp verili durum

$$v(\hat{\kappa}_{BP}) = \frac{v(p_a)}{(1-p_e)^2} \quad (2.52)$$

Burada,

$$v(p_a) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n a_i^2 \quad (2.53)$$

$v(p_a)$: Yüzde uyumun varyansını temsil etmektedir.

f : Eş. 2.13'te tanımlandığı gibidir.

$$a_i = \varepsilon_i / \theta \sum_{k,l}^q w_{kl} (\delta_{kl}^{(i)} - p'_{kl}) \quad (2.54)$$

ε_i : Eş. 2.24'te tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2’de tanımlandığı gibidir.

$\delta_{kl}^{(i)}$: Eş. 2.25’te tanımlandığı gibidir.

p'_{kl} : Eş. 2.6’daki gibidir.

2.4.4. Krippendorff’un α katsayısı

Krippendorff (1970, 1978, 2004) içerik analizi alanında çalışan araştırmacılar tarafından sıklıkla kullanılan, alfa olarak bilinen bir uyum katsayısı önermiştir. Krippendorff’un α ’sının hesaplanması amacıyla, sadece her iki değerlendirici tarafından sınıflanmış denekler kullanılır. Eksik skorlanan denekler tamamen analizden hariç tutulur. Krippendorff α sadece üç veya daha fazla değerlendiricili durumlarda eksik skorlamaya uyum sağlamaktadır.

n , her iki değerlendirici tarafından skorlanan deneklerin sayısı olsun. Krippendorff α aşağıdaki şekilde tanımlanır:

$$\hat{\alpha}_K = \frac{p_a^* - p_e}{1 - p_e}, \quad \begin{cases} p_a^* = (1 - \varepsilon_n)p_a + \varepsilon_n, \varepsilon_n = 1/(2n) \\ p_a = \sum_{k,l}^q w_{kl} p_{kl}, \quad p_e = \sum_{k,l}^q w_{kl} \pi_k \pi_l \end{cases} \quad (2.55)$$

Burada,

w_{kl} : Eş. 2.1 veya Eş. 2.2’de tanımlandığı gibidir.

p_{kl} : Eş. 2.7’de tanımlandığı gibidir.

$$\pi_k = (p_{k+} + p_{+k})/2 \quad (2.56)$$

$$\pi_k = \pi_l \quad (2.57)$$

Yalnızca iki değerlendiriciye dayanan çalışmalar için Krippendorff α ’nın Eş. 2.34’teki ağırlıklandırılmış Scott’un π katsayısı ile çok benzer olduğu gözlenebilir. Kayıp skorlama olmadığında, yakın değerler üretmeleri beklenmektedir.

α varyans kestiricisi

$\hat{\alpha}_K$ ile gösterilen Krippendorff’un α katsayısı yalnızca her iki değerlendirici tarafından skorlanan deneklere dayanır. Kayıp değerli tüm denekler tümüyle analizden çıkarılır.

Krippendorff'un α katsayısının hesaplanması amacıyla, n her zaman her iki değerlendirici tarafından skorlanan denek sayısını temsil eder.

$$v(\hat{\alpha}_K) = \frac{1-f}{n(1-p_e)^2} \{ \sum_{k,l}^q p_{kl} [(1 - \varepsilon_n)w_{kl} - (1 - \hat{\alpha}_K)(\bar{p}_k + \bar{p}_l)]^2 - [(1 - \varepsilon_n)p_a - 2(1 - \hat{\alpha}_K)p_e]^2 \} \quad (2.58)$$

Burada

f : Eş 2.13'te tanımlandığı gibidir.

p_e : Eş. 2.55'te tanımlandığı gibidir.

p_{kl} : Eş. 2.7'de tanımlandığı gibidir.

ε_n : Eş. 2.55'te tanımlandığı gibidir.

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$\hat{\alpha}_K$: Eş. 2.55'te tanımlandığı gibidir.

$$\bar{p}_k = \sum_{l=1}^q w_{kl} \pi_l \quad (2.59)$$

$$\bar{p}_l = \sum_{k=1}^q w_{kl} \pi_k \quad (2.60)$$

p_a : Eş. 2.55'te tanımlandığı gibidir.

Değerlendirici sayısı 2 olduğunda Krippendorff'un α katsayısı kayıp skorlamalarla ilgilenmediğinden yukarıdaki varyans formülü tek formüldür.

Krippendorff'un α katsayısı, kayıp skorlamalar durumunda hesaplanamadığından kayıp veri durumunda bir varyans tahmin edicisine sahip değildir.

2.4.5. Gwet AC₂ katsayısı

$\hat{\kappa}'_G$, aynı zamanda AC₂ olarak anılan, Gwet'in (2008a) AC1 katsayısının ağırlıklandırılmış versiyonudur. Bu katsayı aşağıdaki gibi tanımlanmıştır.

$$\hat{\kappa}'_G = \frac{p_a - p_e}{1 - p_e}, \quad \begin{cases} p_a = \sum_{k,l}^q w_{kl} p'_{kl} \\ p_e = \frac{T_w}{q(q-1)} \sum_{k=1}^q \pi'_k (1 - \pi'_k) \end{cases} \quad (2.61)$$

Burada,

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

p'_{kl} : Eş. 2.6'da tanımlandığı gibidir.

$$\pi'_k = (p'_{k+} + p'_{+k})/2$$

(2.62)

p'_{k+} : Eş. 2.9'da tanımlandığı gibidir.

p'_{+k} : Eş. 2.12'de tanımlandığı gibidir.

AC₂ varyans kestiricisi

$$v(\hat{\kappa}_G) = \frac{1-f}{n(1-p_e)^2} \left\{ \sum_{k,l}^q p_{kl} \left[w_{kl} - 2 \frac{(1-\hat{\kappa}_G)T_w}{q(q-1)} (1 - (\pi_k + \pi_l)/2) \right]^2 - [p_a - 2(1 - \hat{\kappa}_G)p_e]^2 \right\}$$

(2.63)

Burada;

f : Eş 2.16'da tanımlandığı gibidir.

p_e : Eş. 2.61'de tanımlandığı gibidir.

p_{kl} : Eş. 2.7'de tanımlandığı gibidir.

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$\hat{\kappa}_G$: Eş. 2.61'de tanımlandığı gibidir.

$$\pi_k = (p_{k+} + p_{+k})/2$$

(2.64)

p_{k+} : Eş. 2.10'da tanımlandığı gibidir.

p_{+l} : Eş. 2.13'te tanımlandığı gibidir.

$$\pi_l = \pi_k$$

(2.65)

p_a : Eş. 2.58'de tanımlandığı gibidir.

Kayıp verili durum

$$v(\hat{\kappa}_G) = \frac{1-f}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2$$

(2.66)

Burada;

f : Eş 2.16'da tanımlandığı gibidir.

$$u_i = a_i - 2(1 - \hat{\kappa}_G)e_i$$

(2.67)

$$a_i = \frac{\varepsilon_i/\theta}{1-p_e} \sum_{k=1}^q \sum_{l=1}^q w_{kl} (\delta_{kl}^{(i)} - p'_{kl}) \quad (2.68)$$

ε_i : Eş. 2.24'te tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$\delta_{kl}^{(i)}$: Eş. 2.25'te tanımlandığı gibidir.

p'_{kl} : Eş. 2.6'teki tanımlandığı gibidir.

$$e_i = \frac{T_w}{q(q-1)(1-p_e)} \sum_{k=1}^q \pi_k b_k^{(i)} \quad (2.69)$$

p_e : Eş. 2.61'de tanımlandığı gibidir.

π_k : Eş. 2.64'te tanımlandığı gibidir.

$b_k^{(i)}$: Eş. 2.49'da tanımlandığı gibidir.

İki değerlendirici arasındaki uyumu hesaplayan katsayıların hesaplanmasına ilişkin bir uygulama EK-3'te verilmiştir.

2.5. Üç veya Daha Fazla Değerlendiricili Durum İçin Uyum Katsayıları

Sıralı skorlar kullanılarak 2 değerlendirici arasındaki uyumun büyüklüğünü hesaplamak için kullanılan yöntemler, 3 veya daha fazla değerlendirici arasındaki uyumun geniş çaplı bir ölçümünü sağlamamaktadırlar. Üç veya daha fazla değerlendirici için değerlendiriciler-arası uyumu gösteren çeşitli uyum katsayıları aşağıda listelenmiştir.

Conger'in kappa katsayısı (1980)

Ağırlıklandırılmış Brennan-Prediger B-P katsayısı (1981)

Fleiss'in kappa katsayısı (1971)

Krippendorff'un α katsayısı (1970; 1978; 2004)

Gwet'in AC₂ katsayısı (2001)

Bu uyum katsayılarının koşullu çıkarım ve koşulsuz çıkarım durumunda sahip oldukları varyans kestiricileri mevcuttur. Deneklerin rasgele seçilip, değerlendiricilerin rasgele seçilmediği veya ilgilenilen tüm değerlendiricilerin bir değerlendiriciler-arası güvenilirlik

çalışmasına dahil edildiği durumlarda koşullu varyans formülleri kullanılmaktadır. Bu çalışmada, koşullu varyans kestiricilerinin formülleri de verilmiştir.

Değerlendiriciler arası güvenirlik çalışmasında bazı deneklerin her iki değerlendirici tarafından skorlanması aksine tek bir değerlendirici tarafından skorlanması durumunda veri, kayıp skorlamalar içeriyor demektir. Bu çalışmada, ilgili koşullu varyans formülasyonunun yanı sıra, kayıp veri durumunda kullanılabilecek varyans kestirimi de verilmiştir.

g ve h olarak etiketlenen 2 değerlendirici arasındaki şansa bağlı olmayan (şans düzeltmeli) uyum katsayısı genellikle aşağıdaki şekildedir:

$$\hat{\kappa}_{(g,h)} = \frac{p_a(g,h) - p_e(g,h)}{1 - p_e(g,h)} = 1 - \frac{1 - p_a(g,h)}{p_e(g,h)} \quad (2.70)$$

Conger (1980), 3 veya daha fazla değerlendiricili durumda Cohen Kappa'nın doğru genelleştirmesini elde etmek için kullandığı yaklaşım aşağıdaki şekilde tanımlamıştır:

- İlk olarak r tane değerlendiricinin başlangıçtaki kümesi arasından değerlendiricilerin tüm (g,h) çiftleri oluşturulur. Bu şekilde toplam $r(r-1)/2$ tane çift vardır.
- Değerlendiricilerin tüm $r(r-1)/2$ çiftleri boyunca $[1 - p_a(g,h)]$ ve $[1 - p_e(g,h)]$ gibi iki ifadenin ayrı ayrı ortalamaları alınır. Bu işlem, ikisinden biri pay ve diğeri payda olmak üzere iki ortalama üretir.
- Genelleştirilmiş uyum katsayısı daha sonra önceki adımda hesaplanan iki ortalamanın oranının Eş. 2.70'in çifte özel oranı $[1 - p_a(g,h)]/[1 - p_e(g,h)]$ ile değiştirilmesi yoluyla elde edilir.

A, B ve C gibi 3 değerlendiricili bir deney düşünölsün. Değerlendirici sayısı $r=3$ 'tür ve değerlendirici gruplarından $3 \times (3-1)/2 = 3$ değerlendirici çift oluşturulabilir. Bu çiftler (A,B), (A,C) ve (B,C) şeklindedir. Bu nedenle, bu üç çiftin her biri için ayrı ayrı (κ , π veya B-P katsayılarından birini kullanarak) uyum katsayısı hesaplanabilir. Bu, çiftlerin her biri yüzdelik uyum p_a 'yı ve şansa bağlı yüzdelik uyum p_e 'yi üretecektir. Bu her iki nicelik daha sonra yukarıda bahsedilen geniş çaplı uyum katsayısını elde etmek için kullanılabilecektir.

r_{ik} ; i . deneğe belli bir x_k skorunu atayan değerlendiricilerin sayısı olsun(eğer çalışmada sayısal skorlara karşı olarak kategoriler kullanılmışsa bu i . deneği k . kategoriye sınıflayan değerlendiricilerin sayısı olacaktır). i . deneği skorlamış olan değerlendiricilerin sayısı r_i ile gösterilir. Analizde her zaman yalnızca 1 veya daha fazla değerlendirici tarafından skorlanan deneklerin içerileceği varsayılacaktır. Benzer şekilde, q 'nın da bir değerlendirici tarafından kullanılması olası kategori (veya sayısal skorların) sayısını göstermekte olduğu varsayılmıştır. Araştırmacılardan, tüm kategoriler veya skorları iyi seçmeleri ve kullanımları için mantıklı bir şansa sahip olmalarını güvence altına almaları beklenmektedir. Kullanımının muhtemel olmayacağı varsayılan kategoriler herhangi bir değerlendiriciler-arası güvenilirlik çalışmasının parçası olmamalıdır.

2.5.1. Conger'in kappası

Üç ve daha fazla değerlendiricinin arasındaki uyumu değerlendirmek için Conger'in kapa katsayısının formülü aşağıda verilmiştir (Conger, A.J.1980).

$$\hat{\kappa}_C = \frac{p_a - p_e}{1 - p_e} \begin{cases} p_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_i - 1)} \\ p_e = \sum_{k,l} w_{kl}(\bar{p}_{+k}\bar{p}_{+l} - s_{kl}^2/r) \end{cases} \quad (2.71)$$

Burada;

$$r_{ik}^* = \sum_{l=1}^q w_{kl} r_{il} \quad (2.72)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\bar{p}_{+k} = \frac{1}{r} \sum_{g=1}^r p_{gk} \quad (2.73)$$

$$p_{gk} = \frac{n_{gk}}{n_g} \quad (2.74)$$

$$\bar{p}_{+l} = \frac{1}{r} \sum_{g=1}^r p_{gl} \quad (2.75)$$

$$p_{gl} = \frac{n_{gl}}{n_g} \quad (2.76)$$

$$s_{kl}^2 = \frac{1}{r-1} \left(\sum_{g=1}^r p_{gk} p_{gl} - r \times \bar{p}_{+k} \times \bar{p}_{+l} \right) + \quad (2.77)$$

p_{gk} : Eş. 2.74'te tanımlandığı gibidir.

p_{gl} : Eş. 2.76'da tanımlandığı gibidir.

$\bar{p}_{+k}$: Eş. 2.73'te tanımlandığı gibidir.

$\bar{p}_{+l}$: Eş. 2.75'te tanımlandığı gibidir.

Spesifik bir skor x_k için uyum, yalnızca x_k skoru ve i . denekle ilişkilendirilen değerlendiricilerin sayısı olan r_{ik} 'ya bakarak değil ayrıca x_k skoru ile kısmi uyumu temsil eden diğer x_l skorlarına bakılarak da belirlenir. $k \neq l$ olmak üzere x_k ve x_l gibi iki skor w_{kl} ağırlığı sıfırdan farklı olduğunda kısmi uyum içerisindedirler. x_k 'ya nazaran x_k dışındaki diğer skorların uyumunun büyüklüğüne katkısı $l = 1, 2, \dots, q$ ya tüm kategoriler için r_{il} 'lerin ağırlıklandırılmış toplamı r_{ik}^* 'nın değeri tarafından belirlenir.

Conger'in kappası varyans kestiricisi

$$v(\hat{\kappa}_C) = \frac{(1-f)}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\hat{\kappa}_{C|i}^* - \kappa_C)^2 \quad (2.78)$$

Burada,

f : Eş 2.16'da tanımlandığı gibidir.

$$\hat{\kappa}_{C|i}^* = \hat{\kappa}_{C|i} - 2(1 - \kappa_C)(p_{e|i} - p_e)/(1 - p_e) \quad (2.79)$$

$\hat{\kappa}_C$: Eş. 2.71'de tanımlandığı gibidir.

p_e : Eş. 2.71'de tanımlandığı gibidir.

$$\hat{\kappa}_{C|i} = \begin{cases} (n/n')(p_{a|i} - p_e)/(1 - p_e), & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.80)$$

p_e : Eş. 2.71'de tanımlandığı gibidir.

$$p_{a|i} = \begin{cases} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_{i-1})}, & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.81)$$

r_{ik}^* : Eş. 2.69'da tanımlandığı gibidir.

$$p_{e|i} = \frac{1}{r(r-1)} \sum_{g=1}^r \sum_{k=1}^q \bar{w}_{gk}^{(i)} (r\bar{p}_{+k} - p_{gk}) \quad (2.82)$$

$$\bar{w}_{gk}^{(i)} = \sum_{l=1}^q \delta_{gl}^{(i)} (w_{kl} + w_{lk}) / 2 \quad (2.83)$$

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\delta_{gl}^{(i)} = \begin{cases} 1, & g \text{ değerlendiricisi } i. \text{ denegi } l. \text{ kategoriye sınıfladıysa,} \\ 0, & \text{dd} \end{cases} \quad (2.84)$$

$\bar{p}_{+k}$: Eş. 2.73'te tanımlandığı gibidir.

p_{gk} : Eş. 2.74'te tanımlandığı gibidir.

2.5.2. Brennan-Prediger katsayısı

Brennan-Prediger uyum katsayısı aşağıdaki gibi tanımlanmıştır (Brennan-Prediger, 1981).

$$\hat{\kappa}_q = \frac{p_a - p_e}{1 - p_e}, \quad p_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_{i-1})} \quad ve \quad p_e = \frac{1}{q^2} \sum_{k,l} w_{kl} \quad (2.85)$$

Burada;

r_{ik}^* : Eş. 2.72'de tanımlandığı gibidir.

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

B-P katsayısının varyans kestiricisi

$$v(\hat{\kappa}_q) = \frac{(1-f)}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\hat{\kappa}_{q|i} - \hat{\kappa}_q)^2 \quad (2.86)$$

Burada,

f : Eş 2.16'da tanımlandığı gibidir.

$$\hat{\kappa}_{q|i} = \begin{cases} (n/n')(p_{a|i} - p_e)/(1 - p_e), & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.87)$$

$p_{a|i}$: Eş. 2.81'de tanımlandığı gibidir.

p_e : Ş. 2.85'te tanımlandığı gibidir

$\hat{\kappa}_q$: Eş. 2.85'te tanımlandığı gibidir.

2.5.3. Fleiss'in kappa katsayısı

Fleiss Kappa aşağıdaki gibi tanımlanmıştır (Fleiss, 1971).

$$\hat{\kappa}_F = \frac{p_a - p_e}{1 - p_e}, \quad p_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_{i-1})} \quad p_e = \sum_{k,l} w_{kl} \pi_k \pi_l \quad (2.88)$$

r_{ik}^* : Eş. 2.72'de tanımlandığı gibidir.

w_{kl} : Eş 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\pi_k = \frac{1}{n} \sum_{i=1}^n \frac{r_{ik}}{r_i} \quad (2.89)$$

$$\pi_l = \pi_k \quad (2.90)$$

Fleiss kappa katsayısının varyans kestiricisi

$$v(\hat{\kappa}_F) = \frac{(1-f)}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\hat{\kappa}_{F|i}^* - \hat{\kappa}_F)^2 \quad (2.91)$$

Burada,

f : Eş 2.16'da tanımlandığı gibidir.

$$\hat{\kappa}_{F|i}^* = \hat{\kappa}_{F|i} - 2(1 - \kappa_F) \frac{p_{e|i} - p_e}{1 - p_e} \quad (2.92)$$

$$\hat{\kappa}_{F|i} = \begin{cases} ({}^n/n')(p_{a|i} - p_e)/(1 - p_e), & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.93)$$

$p_{a|i}$: Eş. 2.81'de tanımlandığı gibidir.

p_e : Eş. 2.88'de tanımlandığı gibidir.

κ_F : Eş 2.88'de tanımlandığı gibidir.

$$p_{e|i} = \sum_{k=1}^q \bar{\pi}_k r_{ik} / r_i \quad (2.94)$$

$$\bar{\pi}_k = (\bar{\pi}_{k+} + \bar{\pi}_{+k}) / 2 \quad (2.95)$$

$$\bar{\pi}_{k+} = \sum_{l=1}^q w_{kl} \pi_l \quad (2.96)$$

$$\bar{\pi}_{+l} = \sum_{k=1}^q w_{kl} \pi_k \quad (2.97)$$

$$\bar{\pi}_{k+} = \bar{\pi}_{+k} \quad (2.98)$$

2.5.4. Krippendorff'un α katsayısı

Bu bölümde, 2 değerlendiricili durum için Alt bölüm 2.2.4'te verilen Krippendorff α katsayısının çoklu-değerlendiricili duruma genişletilmesi ele alınmıştır. Her ne kadar her iki formülasyonda tamı tamına aynı sonuçları verse de, Gwet (2014:88) tarafından genelleştirilen Krippendorff'un α katsayısının, Krippendorff'un formülasyonundan farklı olduğunu belirtmek gerekmektedir. Gwet (2014:88) tarafından önerilen bu formülasyon, ondan bu tezde daha sonra ele alınacak olan uyum katsayılarının karşılaştırılacağı bölüm için basit bir varyans tahmin edicisi türetmeyi mümkün kılmasının yanında, çok daha basit ve etkin bir hesaplama süreci de sağlamaktadır.

Krippendorff α 'yı hesaplamak maksadıyla, sadece tek bir gözlemci tarafından skorlanan veya hiç skorlanmamış tüm denekler analizden hariç tutulmalıdır.

α katsayısı aşağıdaki gibi tanımlanır:

$$\hat{\alpha}_K = \frac{p_a - p_e}{1 - p_e}, \quad \begin{cases} p_a = (1 - \varepsilon_n)p'_a + \varepsilon_n, \quad \varepsilon_n = 1/(n' \times \bar{r}) \\ p'_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{\bar{r}(r_i - 1)} \\ p_e = \sum_{k,l} w_{kl} \pi_k \pi_l \end{cases} \quad (2.99)$$

Burada,

r_{ik}^* : Eş. 2.69'da tanımlandığı gibidir.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\pi_k = \frac{1}{n'} \sum_{i=1}^{n'} \frac{r_{ik}}{\bar{r}} \quad (2.100)$$

Krippendorff'un α katsayısı aslında daha önceden sunulan Fleiss'in Genelleştirilmiş Kappa'sı $\hat{\kappa}_F$ ya özellikle hiçbir kayıp skorlama yokken çok benzerdir. Her iki katsayıda kayıp skorlama uygulamalarında ve dayandıkları yüzde uyum p_a larda farklılık gösterir.

Krippendorff α katsayısının varyans kestiricisi

$$v(\hat{\alpha}_K) = \frac{(1-f)}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\hat{\alpha}_{K|i}^* - \hat{\alpha}_K)^2 \quad (2.101)$$

Burada,

f : Eş. 2.16'da tanımlandığı gibidir.

$$\hat{\alpha}_{K|i}^* = \hat{\alpha}_{K|i} - (1 - \alpha_K) \frac{p_{e|i} - p_e}{1 - p_e} \quad (2.102)$$

$$\hat{\alpha}_{K|i} = (p_{a\varepsilon_n} - p_e)/(1 - p_e) \quad (2.103)$$

$$p_{a\varepsilon_n} = (1 - \varepsilon_n) \left[p_{a|i} - \frac{p_a(r_i - \bar{r})}{\bar{r}} \right] + \varepsilon_n \quad (2.104)$$

ε_n : Eş. 2.99'da tanımlandığı gibidir.

α_K : Eş. 2.99'da tanımlandığı gibidir.

p_e : Eş. 2.99'da tanımlandığı gibidir.

$$p_{a|i} = \sum_{k=1}^q \frac{r_{ik}(r_{ik}^*-1)}{\bar{r}(r_i-1)} \quad (2.105)$$

r_{ik}^* : Eş. 2.72'de tanımlandığı gibidir.

$$p_a = \frac{1}{n} \sum_{i=1}^n p_{a|i} \quad (2.106)$$

$$p_{e|i} = \sum_{k=1}^q \bar{\pi}_k \frac{r_{ik}}{\bar{r}} - (r_i - \bar{r})/\bar{r} \quad (2.107)$$

$$\bar{\pi}_k = (\bar{\pi}_{k+} + \bar{\pi}_{+k})/2 \quad (2.108)$$

$$\bar{\pi}_{k+} = \sum_{l=1}^q w_{kl} \pi_l \quad (2.109)$$

$$\bar{\pi}_{+l} = \sum_{k=1}^q w_{kl} \pi_k \quad (2.110)$$

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

2.5.5. Gwet AC₂ katsayısı

AC₂ katsayısı Gwet AC₁'in ağırlıklandırılmış versiyonudur (Gwet, 2008a). Aşağıdaki gibi tanımlanır:

$$\hat{K}_G = \frac{p_a - p_e}{1 - p_e}, \quad \begin{cases} p_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^*-1)}{r_i(r_i-1)} \\ p_e = \frac{T_w}{q(q-1)} \sum_{k,l} \pi_k (1 - \pi_k) \end{cases} \quad (2.111)$$

Burada,

r_{ik}^* : Eş. 2.72'de tanımlandığı gibidir.

T_w : Tüm ağırlık değerleri olan w_{kl} 'lerin toplamıdır.

w_{kl} : Eş. 2.1 veya Eş. 2.2'de tanımlandığı gibidir.

$$\pi_k = \frac{1}{n} \sum_{i=1}^n \frac{r_{ik}}{r_i} \quad (2.112)$$

Spesifik bir skor x_k için uyum, sıfırdan farklı her iki skorla ilişkilendirilen w_{kl} ağırlığı bakımından x_k ile kısmi uyumda olan tüm x_l skorlarına bakılarak tanımlanır. x_k ile ilişkili uyumun büyüklüğüne x_k haricindeki diğer skorların katkısı $l; = 1, 2, \dots, q$ gibi tüm kategoriler için r_{il} 'lerin ağırlıklı toplamı olan r_{ik}^* değeriyle tanımlanır.

Gwet AC₂ varyans kestiricisi

$$v(\hat{\kappa}_G) = \frac{(1-f)}{n} \cdot \frac{1}{n-1} \sum_{i=1}^n (\hat{\kappa}_{G|i}^* - \hat{\kappa}_G)^2 \quad (2.113)$$

Burada,

f : Eş 2.16'da tanımlandığı gibidir.

$$\hat{\kappa}_{G|i}^* = \hat{\kappa}_{G|i} - (1 - \hat{\kappa}_G) \frac{p_{e|i} - p_e}{1 - p_e} \quad (2.114)$$

$$\hat{\kappa}_{G|i} = \begin{cases} (n/n')(p_{a|i} - p_e)/(1 - p_e), & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.115)$$

$$p_{a|i} = \begin{cases} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^* - 1)}{r_i(r_i - 1)}, & \text{eğer } r_i \geq 2 \\ 0, & \text{dd} \end{cases} \quad (2.116)$$

p_e : Eş 2.111'de tanımlandığı gibidir.

r_{ik}^* : Eş. 2.72'de tanımlandığı gibidir.

$\hat{\kappa}_G$: Eş 2.111'de tanımlandığı gibidir.

$$p_{e|i} = \frac{T_w}{q(q-1)} \sum_{k=1}^q \frac{r_{ik}}{r_i} (1 - \pi_k) \quad (2.117)$$

π_k : Eş. 2.112'de tanımlandığı gibidir.

Üç veya daha fazla değerlendirici arasındaki uyumu hesaplamak için kullanılan katsayıların hesaplaması EK-4.'te gösterilmiştir.

3. DEĞERLENDİRİCİLER-ARASI GÜVENİLİRLİK KATSAYISINI KİYASLAMA YÖNTEMLERİ

Uyum katsayısının büyüklüğü verildiğinde, değerlendiriciler arasındaki uyum boyutunun “Mükemmel uyum”, “İyi derecede uyum” veya “Zayıf uyum” olarak düşünülebilmesi hiç şüphesiz ki araştırmacıların sonuçları yorumlaması açısından oldukça önemlidir. Önceki bölümde ele alınan ve bir nicelik olarak ifade edilen değerlendiriciler arası güvenilirliği ölçen uyum katsayısını herhangi bir ölçek olmadan uyumun derecesini (“mükemmel”, “iyi”, “zayıf” gibi) yorumlamak güçtür. Literatürde ingilizce karşılığı “Benchmarking” ve Türkçesi “Kıyaslama” olarak geçen bazı karşılaştırma ölçekleri önerilmiştir. Yaygın olarak kullanılan ölçekler Landis-Koch (1977a), Fleiss (1981) ve Altman (1991) ölçekleridir. Ancak elde edilen bir uyum katsayısının bu ölçeklerdeki etiketlerle değerlendirilmesi istatistiksel bir temele dayanmamaktadır. Gwet (2014:177) kümülatif olasılıklar ve standart hata bilgisini de işin içine katarak uyum katsayısının ait olabileceği aralığı hesaplayan yeni bir yöntem önermiştir.

Bu bölümde bu ölçeklerin tanıtılmasının yanı sıra güçlü ve zayıf yönleri de istatistiksel bir çerçevede (güven aralığı, standart hata vb.) değerlendirilmiştir.

3.1. Kıyaslama Kavramı ve Uyumun Büyüklüğü

“Değerlendiriciler arasındaki uyumun büyüklüğü” kavramı, sadece hesaplanan bir uyum katsayısının niceliği ele alındığında değerlendiriciler arasındaki uyumu temsil etme bakımından net bir kavram değildir. Değerlendiriciler-arası güvenilirlik katsayısı mantıksal bir yönde formüle edildiğinde, araştırmacılara, bu muğlak kavramın elle tutulur ve yaygın bir temsilini sağlamaktadır. Bu formülasyonlarda kullanılan birçok farklı mantık uyum katsayılarının çeşitli formlarını ortaya çıkarır ancak bir değerlendiriciler-arası güvenilirlik katsayısının kullanışlı olabilmesi için araştırmacıların elde ettikleri büyüklüğü yorumlayabilmesi gerekmektedir. Her ne kadar hesaplanmış bir uyum katsayısı birbirleri arasında uyum içinde olan değerlendiricilerin arasında bir uyumun olduğunu gösterse de, bu uyumun büyüklüğü hakkında bir ölçek olmaksızın pek de işe yaramayacaktır.

Örneğin 0,5 olarak ölçülen bir uyum katsayısı elde edildiğinde değerlendiriciler arasındaki uyum;

- “İyi”, “ortalama” veya “kötü” olarak mı ele alınmalıdır?
- Bu değer araştırmayı yorumlayabilmek için “kabul edilebilir” bir değer midir?
- Bu gibi bir durumda literatürdeki mevcut kıyaslama ölçekleri kullanılarak bir yorum yapılması istatistiksel olarak ne kadar doğrudur?

Bu soruların yanıtı ilerleyen aşamalarda detaylı olarak verilmiştir.

Değerlendiriciler arası güvenilirlik katsayılarının gelişim sürecinin erken dönemlerinde, araştırmacılar, kestirdikleri uyum katsayısı ile uyum miktarı kavramını ilişkilendirmek için pratik bir yola ihtiyaç duymaktaydılar. Uygulayıcılar, “iyi” olarak dikkate alabilecekleri uyum büyüklüğü ötesinde, Kappa için bir alt-sınır talep etmişlerdi. Tahmin edilen değerlendiriciler-arası güvenilirlik katsayılarının, sonradan geliştirilen uyumun büyüklüğü değerlendirilirken kullanılan ölçeklerden önce, önceden tanımlanan bir alt-sınır ile karşılaştırılmasına literatürde “Kıyaslama-Benchmarking” denir ve alt-sınırlar karşılaştırmayı “Kıyaslama” yapmak için kullanılır.

Birçok bilimsel alanda “kabul edilebilir” olanı “kabul edilemez” olandan ayırt etmek için kalite standartları kullanılır. Her standart kendi kapsamına özgüdür. Tıpkı kalite standartları gibi, kıyaslama ölçekleri de değerlendiriciler arası güvenilirliğin bir standardı gibi düşünülebilir. Bu benzetme aşağıdaki iki sorunun yanıtlanmasını gerektirir:

1. “İyi” olarak düşünülen bir uyum büyüklüğünü iyi yapan şey nedir?
2. Değerlendiriciler-arası güvenilirlik katsayısı, uyum büyüklüğünün iyi olarak düşünülebilmesi için ne kadar yüksek olmalıdır?

Literatürde bu iki soru kappa katsayısı ekseninde yanıtlanmaya çalışılmıştır. İlk olarak Landis ve Koch (1977a) uygulayıcılar arasında geniş ölçüde kullanılan kıyaslama ölçeklerinden birini sunmuşlardır. Kappa istatistiğini uzun bir süre boyunca kullanmış olan araştırmacılar, önerilen bu kıyaslama ölçeğini faydalı bulmuşlardır. Sonrasında, Fleiss (1981) ve Altman (1991) benzer ölçekler geliştirmişlerdir. Gwet (2014: 163) ise mevcut tüm uyum katsayıları ve ölçekler kullanabilecek çok daha sağlam bir kıyaslama metodu geliştirerek, uygulayıcılara istatistiksel olarak daha anlamlı ve güvenilir sonuçlar sunmaktadır. Bu yöntemlerden ilerleyen alt başlıklarda bahsedilmiştir.

3.2. Deneyisel Koşulların Etkisi

Kıyaslama yapmak için şu ana kadar literatürde yayınlanmış olan pek çok çalışmanın pratikte uygulanmasıyla elde edilmiş olan bilgi birikimi oldukça fayda sağlarken, bu uygulamalarda deneyisel koşulların tahmin edilmiş uyum katsayısı üzerindeki etkisini ihmal etmek, bu katsayıların önemini ifade ederken doğru yorumu yapamamakla sonuçlanabilmektedir. Deney koşulları, bir değerlendiriciler-arası güvenilirlik çalışmasına katılan denek, değerlendirici ve kategori sayısı olarak özetlenebilir. Gwet (2014: 169) Monte-Carlo simülasyonu yoluyla kappa ve pi katsayılarının 100000 tesadüfi denemede deneyisel koşullardan nasıl etkilendiğini göstermiştir.

Deney koşullarının etkisini anlatmak amacıyla Gwet (2014;165) şöyle bir örnek vermiştir:

“0,5 gibi bir uyum katsayısı, literatürde bilinen tüm kıyaslama ölçeklerine göre “orta derecede uyum” olarak etiketlenmektedir. Bu değerlendirme 25 denek, 3 değerlendirici ve 4 yanıt kategorisini içeren bir çalışma için kabul edilebilir olabilir fakat bu büyüklükteki bir uyum katsayısı eğer çalışma 10 denek, 2 değerlendirici ve 2 yanıt kategorisine dayanıyorsa istatistiksel olarak önemli bile değildir.”

İstatistiksel önem, gerçekteki katsayının kestirilen değerinin yansıttığı bilginin doğruluğunu sınamak için gereklidir. Gerçekte 0 değerine sahip bir güvenilirlik katsayısı, örnekleme hataları ile birlikte 0,5 kadar yüksek bir değerde kestirilebilir. Ek olarak, örnekleme hatalarından bağımsız olarak sadece deneyisel koşulların etkisi ile 0,5’ten büyük bir kestirim bile yapılabilmektedir. Sadece deneydeki örnekten elde edilmiş bir katsayının istatistiksel öneme sahip olup olmadığı değerlendirmeden bir nicelik olarak tek başına değerlendirilmesi bile uyumun mantıksal boyutunun zayıf olabileceğini göstermektedir.

Kıyaslama, bir değerlendiriciler-arası güvenilirlik çalışmasının sonuçlarının geniş bir kitleye tebliğ edilmesi ve uygulayıcılara uyum istatistiklerinin ne anlattığını yorumlayabilmeleri konusunda yardım etmektedir. Her ne kadar, yayınlandıkları makalelerde bu ölçekler kappa istatistiğini ele almış olsa da, pratikte diğer uyum katsayıları ile birlikte kullanılmamalarını tavsiye eden herhangi bir çalışma bulunmamaktadır.

Bu tezde, 2. Bölümde verilen katsayıların uygulamasını sağlamak amacıyla gerçek veriler kullanılmış olup, deney koşullarının etkisi simülasyon çalışmasıyla gösterilmemiştir.

3.3. Literatürdeki kıyaslama yöntemleri

3.3.1. Landis ve Koch kıyaslama ölçeği

Landis ve Koch (1977a) iki nörolog arasındaki uyumun büyüklüğünü analiz etmek için, çoklu skleroz teşhisine ilişkin bir değerlendiriciler-arası güvenilirlik verisinde uyum istatistiklerinden kappa katsayısını kullanmışlardır. Bu çalışmada metodolojilerine uygun hesaplamalar yaptıktan sonra, kendilerinin tanımıyla “açıkça keyfi” olarak aralıklar belirlemişlerdir. Bu aralıkların çalışma verisi için yararlı “kıyaslama aralıkları” olduğunu sağladığını da iddia etmişlerdir. Ölçek 6 farklı kıyas aralığından oluştuğu için dar aralıklarda özgün değerlendirmelere olanak vermektedir. Bu, ölçeğin bir avantajı olarak ifade edilebilir. Uyum katsayısına ilişkin kıyas aralıkları aşağıdadır:

Çizelge 3.1. Landis ve Koch (1977a) kappa için belirlediği kıyaslama ölçeği

Kappa İstatistiği	Uyumun Gücü
<0,00	Yetersiz
0,0 – 0,20	Önemsiz
0,21 – 0,40	Zayıf
0,41 – 0,60	Orta
0,61 – 0,80	Önemli
0,81 – 1,00	Neredeyse Mükemmel

Bu ölçeğe göre, örneğin 0,65 olarak hesaplanan bir kappa katsayısı r tane değerlendirici arasındaki uyum büyüklüğünün “Önemli” olduğunu anlatır. Daha öncede belirtildiği gibi, 0,65 gibi bir değer, istatistiksel önemine bakılmaksızın, bu ölçeğe göre “önemli” derecede bir uyumu tespit ettiğini söylemek istatistiksel olarak hatalı bir yorum yapılmasına neden olabilir. Bu ölçek, Everitt (1992) gibi diğer yazarlar tarafından desteklenmiştir.

3.3.2. Fleiss kıyaslama ölçeği

Fleiss (1981), Landis ve Koch Kıyaslama Ölçeğinin ilk 3 değer aralığını tek bir değer aralığına indirgeyerek yeni bir kıyaslama ölçeği geliştirmiştir. Bu ölçek, 0,40’tan küçük

hesaplanan uyum katsayılarının “Zayıf” derecede uyum gösterdiğini belirten oldukça geniş kıyaslama aralıklarına sahip bir ölçektir ki, bu özellik, ölçeğin en büyük dezavantajları arasında yer alır. Uyum katsayısına ilişkin kıyas aralıkları aşağıdadır:

Çizelge 3.2. Fleiss (1981) kappa için belirlediği kıyaslama ölçeği

Kappa İstatistiği	Uyumun Gücü
<0,40	Zayıf
0,40 – 0,75	Ortalamadan iyiye doğru
0,75<	Mükemmel

Landis ve Koch Kıyaslama Ölçeğinde 0,75 gibi bir değer “Önemli” derecede uyum var olarak değerlendirilebiliyorken, Fleiss Kıyaslama Ölçeğinde “Ortalamadan iyiye doğru” olarak değerlendirilir. Oluşan bu algı farklılığı, açıkça, Fleiss’in kıyaslama ölçeğinde oldukça geniş aralıklar kullanmasından kaynaklanmaktadır. Ayrıca “Ortalamadan iyiye doğru” nitelendirmesi, pek çok kişi için, farklı algılanmaya imkân verecek kadar da muğlak bir ifadedir.

3.3.3. Altman kıyaslama ölçeği

Altman (1991), Landis ve Koch (1977) tarafından önerilen ölçeği geliştirmiştir. Landis ve Koch Kıyaslama Ölçeği ile Altman’ın Kıyaslama Ölçeği arasındaki farklar,

- Landis ve Koch Kıyaslama Ölçeğinin ilk iki değer aralığının Altman tarafından birleştirilmiş olup “Zayıf” uyum olarak etiketlenmiş olmasıdır.
- “Önemli” olan uyum etiketi “İyi”, “Neredeyse Mükemmel” olan uyum etiketi ise “Çok iyi olarak değiştirilmiştir.

Uyum katsayısına ilişkin kıyas aralıkları aşağıdadır:

Çizelge 3.3. Altman(1991) kappa için belirlediği kıyaslama ölçeği

Kappa İstatistiği	Uyumun Gücü
<0,20	Önemsiz
0,21 – 0,40	Zayıf
0,41 – 0,60	Orta
0,61 – 0,80	İyi
0,81 – 1,00	Çok iyi

Şu ana kadar bahsi geçen kıyaslama ölçeklerinin herhangi biri için çok iyi veya çok kötü kararını vermek şüphesiz ki hatalı olabilir. Bunun yerine, dezavantajları ve avantajlarını belirtmek çok daha uygun olacaktır.

3.3.4. Gwet'in kıyaslama modeli

Gwet (2014) literatürdeki herhangi bir uyum katsayısı ile hesaplanmış bir uyum niceliğinin yine literatürdeki herhangi bir kıyaslama ölçeği kullanılarak değerlendirilebilmesi için, istatistiksel olarak daha anlamlı bir kıyaslama modeli önermiştir. Model kümülatif olasılık hesabına dayanarak, güven aralığı bilgisini kullandırmaktadır. Bu haliyle, literatürdeki diğer kıyaslama ölçeklerinden daha fazla istatistiksel bir çerçevede değerlendirme imkânı sunar. Gwet (2014:173) geliştirdiği yöntemle ilgili şu sözleri paylaşmıştır:

“...bir uyum katsayısının uygun kıyaslaması onun doğruluğunu ve gerçekteki değerini çevreleyen istatistiksel hatalardan kaynaklı belirsizliği dikkate almalıdır. Örneğin, hata payı $\pm 0,05$ olan 0,5 değerli bir uyum katsayısı bize değerlendiriciler arasındaki uyumun gerçek büyüklüğü hakkında hata payı $\pm 0,3$ olup aşırı derecede yüksek hata payıyla yine 0,5 değerli bir uyum katsayısından çok daha güçlü bir hikâye anlatır. Zaten aslına bakıldığında, tek başına hesapladığımız uyum katsayısını kullanarak “bizim kappa’mız Landis-Koch kıyaslama ölçeğine göre “iyi” gibi bir tanımlayıcı sonuç çizemeyiz.... Deneysel veriden hesaplanan uyum katsayısı skorumla sürecine dâhil olan denek ve değerlendiriciler ile birlikte değişen tesadüfi bir sayıdır. Bu nedenle, değerlendiriciler arasındaki uyumu ölçmek için kullanımı doğal olarak olasılıksal olmalıdır...”

Gwet'in önerdiği yöntem

Gwet (2014:174) 3 adımlı bir kıyaslama süreci önermiştir:

1. İlk adım, seçilen belli bir uyum katsayısı ve kıyaslama ölçeği için, kıyaslama seviyelerini tanımlayan aralıkların her birinin içine düşecek olan uyum büyüklüğü olan olasılığı hesaplamayı içerir. Her adım için aitlik olasılığı³ hesaplanır. Bir sonraki alt bölümde bu olabirlik olasılıklarının nasıl hesaplanacağı gösterilmiştir.

³ Aitlik olasılığı, bir uyum katsayısının, farklı uyum seviyeleri için etiketlenen ilişik olduğu herhangi bir kategoriye ait olma olasılığı anlamına gelir.

2. İkinci adım olarak, tepeden (en üst aralık) başlayıp, aşağı doğru inen her aralık için kümülatif olasılıklar hesaplanır. Şöyle ki, en üst aralıkla ilişkilendirilen kümülatif olasılık, i. adımda aynı aralık için hesaplanan olasılıkla aynıdır. Azalan sıradaki ikinci aralık için, kümülatif olasılık i. adımda üstteki iki aralık için hesaplanan olasılıkların toplamıdır.
3. Son olarak, en küçük kümülatif olasılıkla ilişkilendirilen aralıkla tanımlanan kıyaslama seviyesi %95'i aşmaktadır. Bu durumda, %95 veya daha fazla kesinlikle değerlendiriciler arası güvenilirliğin örneğin "Mükemmel" olduğunu sonuçlandıracak bir pozisyonda olunacaktır.

Bu yaklaşım, her ne kadar farklı deney dizaynlara dayansa da, aynı uyum istatistiğini kullanan farklı çalışmalar arasında adil bir kıyaslama olanağı sunacaktır.

Gwet'in Önerdiği Yöntemin Yorumlanması ve Karşılaştırma Olasılıkları

Aşağıdaki adımlar sağlanarak uygulanan yeni kıyaslama yöntemini özetlenebilir (Gwet, 2014:177).

Kestirim aşaması

1. İlgilenilen ağırlıklandırılmış ya da ağırlıklandırılmamış bir uyum istatistiği seçilir (AC_2 , κ , α , B-P, ...).
2. Gözlenen skorlar kullanılarak Adım 1'de seçilen uyum katsayısı(COEFF) hesaplanır.

Kıyaslama olasılıkları aşaması

3. Kullanılacak olan kıyaslama ölçeği seçilir (Landis-Koch, Fleiss veya Altman).
4. Seçilen uyum katsayısına ilişkin standart hata (STDHATA) hesaplanır.
5. Bir kıyaslama seviyesini tanımlayan her bir (a,b) aralığı için IMP-Aralık Aidiyet Olasılığı hesaplanır. Bu, standart Normal değişken Z'nin aralıkta $[(KATSAYI-b)/SE, (KATSAYI-a)/SE]$ yer alma olasılığıdır.

$$IMP = P\left(\frac{KATSAYI-b}{STDHATA} \leq Z \leq \frac{KATSAYI-a}{STDHATA}\right) \quad (3.1)$$

Kümülatif olasılıklar

6. En üstteki kıyaslama seviyesinden başlayıp, en alttaki kıyaslama seviyesine inerek kümülatif kıyaslama olasılıkları hesaplanır.

Final uyum seviyesi

7. Nihai uyum 0,95'i aşan en küçük olasılık değerinin olduğu en yüksek kıyaslama seviyesine karşılık gelmektedir.

AC₁ uyum katsayısının kestiriminin $\hat{\kappa}_G$, yığın parametresi κ_G ve standart hatasının $\sigma_{\hat{\kappa}_G}$ olduğu varsayılın. Gwet (2008b), $\hat{\kappa}_G - \kappa_G / \sigma_{\hat{\kappa}_G}$ oranının standart normal dağılıma uyduğunu göstermiştir.

Sonuç olarak, eğer herhangi bir a değeri için $\hat{\kappa}_G - a / \sigma_{\hat{\kappa}_G}$ oranını aşan standart normal değişken Z için olasılık çok küçükse o zaman bu yığın parametresi κ_G 'nin a 'nın altında olmasının olası olduğuna dair güçlü bir kanıt olarak görülebilir. Bu kıyaslama olasılıklarını değerlendirmek için kullanılan bir yaklaşımdır.

Kıyaslama için olasılıkları kullanma yaklaşımı istatistiksel olarak daha güvenilir bir yaklaşımdır. Gwet (2014:177)'deki yaklaşımını göstermek ve diğer ölçeklerle kıyaslamak adına aşağıda örnek bir uygulama hazırlanmıştır.

Gwet (2014:177) kıyaslama yöntemi, literatürdeki diğer kıyaslama ölçekleri gibi bir kıyaslama ölçeği olmayıp, uyum katsayısının uyum derecesini, seçilen herhangi bir kıyaslama ölçeğine göre istatistiksel olarak daha sağlam bir şekilde değerlendirmeyi sağlayan bir yöntemdir.

Örnek uygulama

Landis ve Koch (1977b) servikte rahim karsinoma in situ evreleri sınıflamasındaki değişimlerin incelenmesi için 7 patolog tarafından 118 örneğin,

- (1) Negatif
- (2) Atipik skuamöz hiperplazi
- (3) Karsinoma in situ
- (4) Erken Stromal İnvazyonla birlikte skuamöz kanser
- (5) İstilacı karsinoma

teşhis kategorisinden birine sınıflandırıldığı hipotetik bir değerlendiriciler arası güvenilirlik deneyini kullanılmıştır. Deney verisi, EK-5'te verilmiştir. Örnek uygulamada bu veriye ilişkin 7 patoloğun sınıflama verisi alınarak ağırlıklandırılmış uyum katsayıları bakımından aşağıdaki gibi hesaplanmıştır.

Çizelge 3.4. Landis ve Koch (1977b) verisine ilişkin hesaplanan değerlendiriciler-arası güvenilirlik katsayıları

YÖNTEM	Katsayı Değeri	Std Hata	95% G.A.
Conger κ	0,601	0,039	0,525-0,677
Gwet'AC ₂	0,811	0,017	0,778-0,844
Fleiss κ	0,595	0,040	0,516-0,674
Krippendorff α	0,596	0,033	0,531-0,660
Brenann-Prediger B-P	0,761	0,018	0,725-0,797
Yüzde Uyum p_a	0,933	0,005	0,923-0,943

Birbirinden farklı şansa bağlı olmayan 5 uyum katsayısının 0,595'ten 0,811'e kadar değiştiği ortaya çıkmıştır. Çizelge 3.4'te her bir uyum katsayısına ait denek bazında standart hata ve güven aralığı da hesaplanmıştır. Katsayı değerleri, bu veri seti 7 patalog içerdiğinden Alt Bölüm 2.5'te verilen katsayı formüllerinden ve standart hatalar da yine bu formüllerle birlikte verilen varyans formüllerinden elde edilmiştir. Güven aralığı ise, *katsayı değeri* $\pm$ (*standart hata* $\times$ 1,96) hesaplamasıyla elde edilmiştir. Standart hata, bahsi geçen herhangi bir katsayı değerinin ortalamadan ne kadar sapma yapmasının beklenebileceğini anlatan basit bir sayıdır. Bu standart hata normal dağılım Z'nin $\alpha=0,05$ 'teki tablo değeri olan 1,96 ile çarpıldığında, bu çarpım, ilgili katsayıyla ilişkilendirilen ve %95 hata payı olarak bilinen hata payını temsil eder. Elde edilen katsayı değerine bu hata payı çıkartıldığında güven aralığının alt sınırı, eklendiğinde ise güven aralığının üst sınırı bulunmuş olur. Bu sonuçlar EK-5'teki denek grubuna bağlı sonuçlardır. Hatırlanmalıdır ki farklı denek grupları farklı sonuçlar üretecektir. Öncelikle elde edilen katsayılar hem Landis-Koch hem Fleiss, hem de Altman'ın Kıyaslama Ölçekleri kullanılarak değerlendirilecektir.

Çizelge 3.5. Landis ve Koch (1977b) verisine ilişkin hesaplanan değerlendiriciler-arası güvenilirlik katsayılarının Landis&Koch, Fleiss ve Altman ölçeklerine göre kıyaslama değerleri

YÖNTEM	Katsayı Değeri	Landis&Koch	Fleiss	Altman
Conger κ	0,601	Orta	Ortalamadan iyiye	Orta
Gwet AC ₂	0,811	Neredeyse Mükemmel	Mükemmel	Çok iyi
Fleiss κ	0,595	Orta	Ortalamadan iyiye	Orta
Krippendorff α	0,596	Orta	Ortalamadan iyiye	Orta
Brenann-Prediger B-P	0,761	Önemli	Mükemmel	İyi

Gwet (2014:177) kıyaslama yönteminin kullanılabilmesi için bu aşamadan sonra aralık aidiyet olasılıklarının hesaplanması gerekmektedir.

Gwet Kıyaslama Yöntemi Hesaplaması (Landis&Koch Ölçeğine göre)

Çizelge 3.6. Uyum katsayılarına göre Landis&Koch kıyaslama aralıklarına ait aitlik olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC ₂	Fleiss κ	α	B-P
0,81-1	Neredeyse Mükemmel	0,00	0,75	0,00	0,00	0,02
0,61-0,80	Önemli	0,51	0,25	0,45	0,45	0,98
0,41-0,60	Orta	0,49	0,00	0,55	0,55	0,00
0,21-0,40	Zayıf	0,00	0,00	0,00	0,00	0,00
0-0,20	Önemsiz	0	0	0	0	0
<0,0	Yetersiz	0	0	0	0	0

Çizelge 3.6, Landis ve Koch (1977b) verisine göre Landis&Koch kıyaslama ölçeğini kullanan beş uyum katsayısının her bir aralığa aitlik olasılıklarını göstermektedir.

Çizelge 3.7. Uyum katsayılarına göre Landis&Koch kıyaslama aralıklarına ait aitlik kümülatif olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC ₂	Fleiss κ	α	B-P
0,81-1,00	Neredeyse Mükemmel	0,00	0,75	0,00	0,00	0,02
0,61-0,80	Önemli	0,51	1,00	0,45	0,45	1,00
0,41-0,60	Orta	1,00	1,00	1,00	1,00	1,00
0,21-0,40	Zayıf	1,00	1,00	1,00	1,00	1,00

Çizelge 3.7. (devam) Uyum katsayılarına göre Landis&Koch kıyaslama aralıklarına ait aitlik kümülatif olasılıkları

0,20-0,00	Önemsiz	1,00	1,00	1,00	1,00	1,00
<0.00	Yetersiz	1,00	1,00	1,00	1,00	1,00

Çizelge 3.7, Landis ve Koch (1977b) verisine göre Landis&Koch kıyaslama ölçeğini kullanan beş uyum katsayısının kümülatif olasılıklarını göstermektedir. Örneğin, Conger κ katsayısına bakılarak 7 patoloğun arasındaki uyumun büyüklüğünün 0 olasılıkla “Çok iyi” olduğu iddia edilebilir. Fakat $\%100(0,00+0,51+0,49=1)$ kesinlikle 7 patoloğun arasındaki uyumun orta derecede bir uyum gösterdiği söylenebilir. Kümülatif olasılıklar, en yüksek aralığa ait olasılık değerinden başlayarak sırasıyla en düşük aralığa kadar mevcut aitlik olasılıklarının bir toplamını ifade etmektedir. Bu nedenle, tek bir karalığa ait olasılık değeri ile yorum yamaktan ziyade, toplam olasılıklarla yorum yapmak daha güvenilir bir temsil sunar. Conger’in κ katsayısının $\%100$ kesinlikle orta bir uyumu temsil ettiği bu şekilde daha büyük bir kesinlikle iddia edilebilir. Gwet (2014:177) önerdiği kıyaslama yöntemi kümülatif olasılıklar için durma noktasını $\%95$ olarak ayarlamıştır. Bu “ $\%95$ ” kuralı, belli bir uyum katsayısına atanan nihai kıyaslama seviyesinin, $\%95$ ’i aşan kümülatif olasılıkla ilişkilendirilen ölçeğin en yüksek seviyesi olmasını şart koşar.

Çizelge 3.7’deki kümülatif olasılık değerleri elde edildikten sonra literatürdeki 3 kıyaslama ölçeğine göre ölçülen uyum seviyeleri Çizelge 3.8’de verilmiştir.

Çizelge 3.8. Gwet’in yeni yöntemi altında kıyaslama

	<u>Kıyaslama Ölçeği</u>		
	Landis&Koch	Fleiss	Altman
Conger κ	Orta	Ortalamadan iyiye	Orta
Gwet AC ₂	Önemli	Ortalamadan iyiye ve Mükemmel	İyi
Fleiss κ	Orta	Ortalamadan iyiye	Orta
α	Orta	Ortalamadan iyiye	Orta
B-P	Önemli	Ortalamadan iyiye ve Mükemmel	İyi

Gwet kıyaslama yöntemi hesaplaması (Fleiss ölçeğine göre)

Çizelge 3.9. Uyum katsayılarına göre Fleiss kıyaslama aralıklarına ait aitlik olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC_2	Fleiss κ	α	B-P
>0,75	Mükemmel	0,00	1,00	0,00	0,00	0,73
0,41-0,75	Ortalamadan iyiye	1,00	0,00	1,00	1,00	0,27
<0,40	Zayıf	0,00	0,00	0,00	0,00	0,00

Çizelge 3.9, Landis ve Koch (1977b) verisine göre Landis&Koch kıyaslama ölçeğini kullanan beş uyum katsayısının her bir aralığa aitlik olasılıklarını göstermektedir.

Çizelge 3.10. Uyum katsayılarına göre Fleiss kıyaslama aralıklarına ait aitlik kümülatif olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC_2	Fleiss κ	α	B-P
>0,75	Mükemmel	0,00	1,00	0,00	0,00	0,73
0,41-0,75	Ortalamadan iyiye	1,00	1,00	1,00	1,00	1,00
<0,40	Zayıf	1,00	1,00	1,00	1,00	1,00

Çizelge 3.10'daki kümülatif olasılık değerleri elde edildikten sonra literatürdeki 3 kıyaslama ölçeğine göre ölçülen uyum seviyeleri Çizelge 3.8'de verilmiştir.

Çizelge 3.11. Gwet'in yeni yöntemi altında kıyaslama

<u>Kıyaslama Ölçeği</u>			
	Landis&Koch	Fleiss	Altman
Conger κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
Gwet AC_2	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi
Fleiss κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
α	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
B-P	Orta-Önemli	Ortalamadan iyiye	Orta-İyi

Gwet kıyaslama yöntemi hesaplaması (Altman ölçeğine göre)

Çizelge 3.12. Uyum katsayılarına göre Altman kıyaslama aralıklarına ait aitlik olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC ₂	Fleiss κ	α	B-P
0,81-1	Çok iyi	0,00	0,75	0,00	0,00	0,02
0,61-0,80	İyi	0,51	0,25	0,45	0,45	0,98
0,41-0,60	Orta	0,49	0,00	0,55	0,55	0,00
0,21-0,40	Zayıf	0,00	0,00	0,00	0,00	0,00
0-0,20	Önemsiz	0,00	0,00	0,00	0,00	0,00

Çizelge 3.12, Landis ve Koch(1977b) verisine göre Altman kıyaslama ölçeğini kullanan beş uyum katsayısının her bir aralığa aitlik olasılıklarını göstermektedir.

Çizelge 3.13. Uyum katsayılarına göre Altman kıyaslama aralıklarına ait aitlik kümülatif olasılıkları

Kıyaslama Aralığı	Tanım	Conger κ	Gwet AC ₂	Fleiss κ	α	B-P
0,81-1,00	Çok iyi	0,00	0,75	0,00	0,00	0,02
0,61-0,80	İyi	0,51	1,00	0,45	0,45	1,00
0,41-0,60	Orta	1,00	1,00	1,00	1,00	1,00
0,21-0,40	Zayıf	1,00	1,00	1,00	1,00	1,00
0,0-0,00	Önemsiz	1,00	1,00	1,00	1,00	1,00

Çizelge 3.13’deki kümülatif olasılık değerleri elde edildikten sonra literatürdeki 3 kıyaslama ölçeğine göre ölçülen uyum seviyeleri Çizelge 3.14’de verilmiştir.

Çizelge 3.14. Gwet’in yeni yöntemi altında kıyaslama

	<u>Kıyaslama Ölçeği</u>		
	Landis&Koch	Fleiss	Altman
Conger κ	Orta	Ortalamadan iyiye	Orta
Gwet AC ₂	Önemli	Ortalamadan iyiye ve Mükemmel	İyi
Fleiss κ	Orta	Ortalamadan iyiye	Orta
α	Orta	Ortalamadan iyiye	Orta
B-P	Önemli	Ortalamadan iyiye ve Mükemmel	İyi

Gwet yöntemi ve literatürdeki mevcut ölçekler kullanılarak belirlenen uyum seviyeleri bu şekildedir.

Çizelge 3.15. Kıyaslama Ölçekleri Bazında Karşılaştırma Çizelgesi

Kıyaslama Ölçekleri	Avantajları	Dezavantajları
Landis ve Koch Kıyaslama Ölçeği	6 farklı aralığa sahiptir. Bu da aralıkların dar olmasından ötürü, katsayısının daha spesifik bir kategoride değerlendirilmesini sağlar.	Aralıklar keyfi seçilmiş
Fleiss Kıyaslama Ölçeği	Uyumu yalnızca mükemmel veya zayıf olup olmadığını bilmek isteyenler için pratiktir.	- Aralıklar oldukça geniş olduğundan diğer ölçeklere göre önemli değerlendirilebilecek bir uyumu, yalnızca ortalamadan iyiye olarak yorumlanmasını sağlayabilir. - Ortalamadan iyiye etiketi muğlak bir etiket olarak değerlendirilebilir.
Altman Kıyaslama Ölçeği	5 farklı aralığa sahiptir. Bu da aralıkların dar olmasından ötürü, katsayısının Fleiss'in kıyaslama ölçeğine göre daha spesifik bir kategoride değerlendirilmesini sağlar.	- Landis ve Koch Kıyaslama Ölçeğinin ilk iki değer aralığının Altman tarafından birleştirilmiş olup "Zayıf" uyum olarak etiketlenmiş olmasıdır. "Önemli" olan uyum etiketi "İyi", "Neredeyse Mükemmel" olan uyum etiketi ise "Çok iyi" olarak değiştirilmiştir.
Gwet Yöntemi	- Standart hata bilgisini kullandığından, direk bir ölçek kullanıp sonuca varmaktan ziyade istatistiksel olarak daha güvenilirlerdir. - Değişik deney tasarımına sahip araştırmaların uyum katsayısı sonuçları için adil bir kıyaslama olanağı sunar.	Standart hata bilgisi ile kümülatif olasılıklar hesaplanarak bulunan aitik aralıkları, genellikle literatürdeki ölçekler direk kullanıldığına elde edilen uyum seviyesinden bir seviye daha düşüktür.

Yukarıdaki Çizelgeye göre, örneğin, hesapladığı uyum katsayısının yalnızca mükemmel ya da zayıf olup olmadığını öğrenmek isteyen bir araştırmacı Fleiss'in Kıyaslama Ölçeğini veya

yeteri kadar iyi bir kategorizasyon arayanlar Landis ve Koch Kıyaslama Ölçeğini tercih edebilirler. Daha kritik bir mesele şudur ki, hangi ölçek kullanılırsa kullanılsın, elde edilen uyum katsayılarının yalnızca bu ölçeklere göre uyumun büyüklüğünü değerlendirmesi hatalı olacaktır. Önerilen yaklaşım, bu uyum katsayılarının hata paylarının incelenmesidir.

Sonuç olarak, denek, değerlendirici ve kategori sayısı, örnek-tabanlı uyum katsayılarını yorumlarken kritik bir hal almaktadır. Bu faktörlerin etkisi uyum katsayısı olasılık dağılımı hesaplanarak gösterilmiştir.

4. BULGULAR

Bu bölümde çeşitli değerlendiriciler arası güvenilirlik çalışmaları için önceki bölümlerde verilen katsayıların hesaplaması gerçekleştirilmiştir. Hesaplama yapılan verilerin kaynakları her analiz öncesi açıklanmış olup, bu veriler AgreeStat 2013,3 Yazılımı yardımıyla hesaplanmıştır.

Değerlendiriciler arası güvenilirlik çalışmalarında temelde merak edilen gerçekteki uyumun sıfırdan farklı olup olmadığıdır. Buna göre, araştırmada merak edilen yığın parametresi olan “gerçekteki uyum”un sıfırdan farklı olup olmadığı da hesaplanan p-değerleri yardımıyla test edilebilir. Tüm katsayılar için öncelikle p-değerlerine bakılarak gerçekteki uyumun sıfırdan farklı olup olmadığı aşağıdaki hipoteze göre test edilmiştir.

$$H_0: \kappa = 0$$

$$H_1: \kappa \neq 0$$

1. tip hata α değeri 0,05 alınmıştır. Çift yanlı bu hipotez testi için $\alpha/2=0,025$ olur ve tüm katsayılar için p-değeri $<0,025$ olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Bulgular değerlendirilirken, H_0 hipotezini hakkında verilen karar buna göre verilmiştir.

Bulguların hesaplanmasında kullanılan kuadratik ağırlıklar Eş. 2.1’e göre hesaplanırken, basit sıralı ağırlıklar Eş. 2.2’ye göre hesaplanmıştır.

4.1. İki Değerlendiricili Durum

Bu bölümde, sıralı 3, 4, 5, 6 ve 7 kategorili gerçek hayattan alınmış veriler, Alt Bölüm 2.5’teki uyum katsayıları hesaplanarak analiz edilmiştir.

4.1.1. Üç kategorili durum

Walsh ve diğerleri (2014), 24 aydan küçük ateşli bebeklerin kliniksel görünüşünün değerlendiriciler arası güvenilirliğini Cohen’in ağırlıklandırılmış κ ’sı, Scott’ın π ’si ve Gwet AC₂’si ile ölçmüşlerdir. Çalışmada 173 kayıtlı hastanın 159’u analiz edilmiştir. Acile yüksek ateş şikayeti ile gelen 159 bebeği iki değerlendirici muayene etmiş ve sırasıyla “Hastalık

belirtisi yok”, “Belirsiz” ve “Hastalık belirtisi” kategorilerine sınıflandırmışlardır. Kategorilerin arasında sıralı bir ilişki olduğu söylenebilir. Bu nedenle “Hastalık belirtisi yok” kategorisi 1 “Belirsiz” kategorisi 2 ve “Hastalık belirtisi” kategorisi 3 rakamıyla etiketlenebilir ve aradaki farklar anlamlı bulunduğu bu sınıflama aralıklı ölçek olarak bile düşünülebilir. Çalışma verisi Çizelge 4.1’deki gibidir:

Çizelge 4.1. Deneklerin değerlendirici ve kategori bazında dağılımı

Değerlendirici A	Değerlendirici B			Toplam	Yüzde
	Hastalık belirtisi yok	Belirsiz	Hastalık Belirtisi		
Hastalık belirtisi yok	103	6	14	123	[77,4%]
Belirsiz	8	0	1	9	[5,7%]
Hastalık Belirtisi	14	2	11	27	[17%]
Toplam	125	8	26	159	[100%]
Yüzde	[78,6%]	[5%]	[16,4%]	[100%]	

Çizelge 4.1’deki dağılım yüzdelerinden A ve B değerlendiricilerinin skorlamaları arasında çok da fazla bir fark olmadığı anlaşılabılır ancak yüzdeler tek başına bir gösterge değildir. Bu nedenle, Çizelge 4.1’deki veriler sıralı ölçeklerde 2 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, Çizelge 4.3’te sonuçlar verilmiştir.

Çizelge 4.2. Çizelge 4.1’in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar

	Kuadratik Ağırlıklar			Basit Sıralı Ağırlıklar		
	1	2	3	1	2	3
1	1	0,75	0	1	0,667	0
2	0,75	1	0,75	0,667	1	0,667
3	0	0,75	1	0	0,667	1

1.Hastalık belirtisi yok

2.Belirsiz

3.Hastalık Belirtisi

Çizelge 4.3. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

KUADRATİK	Uyum Katsayıları	Hesaplanan Katsayı Değeri	Std Hata	95% Güven Aralığı	p-Değeri
	Cohen κ	0,289	0,092	0,106 - 0,472	$2,106 \times 10^{-3}$
	Gwet AC2	0,683	0,056	0,571 - 0,794	$2,316 \times 10^{-24}$
	Scott π	0,289	0,092	0,106 - 0,472	$2,117 \times 10^{-3}$

Çizelge 4.3. (devam) Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

	Krippendorff' α	0,291	0,092	0,109 - 0,473	$1,897 \times 10^{-3}$
	Brenann-Prediger B-P	0,392	0,090	0,215 - 0,568	$2,22 \times 10^{-5}$
	Gözlenen uyum p_a	0,797	0,030	0,738 - 0,856	$1,83 \times 10^{-60}$
BASİT SIRALI	Cohen κ	0,279	0,090	0,101 - 0,457	$2,344 \times 10^{-3}$
	Gwet AC2	0,679	0,055	0,571 - 0,787	$3,91 \times 10^{-25}$
	Scott π	0,279	0,090	0,101 - 0,457	$2,357 \times 10^{-3}$
	Krippendorff' α	0,281	0,090	0,104 - 0,459	$2,109 \times 10^{-3}$
	Brenann-Prediger B-P	0,428	0,081	0,268 - 0,588	$4,08 \times 10^{-7}$
	Gözlenen uyum p_a	0,788	0,030	0,729 - 0,848	$1,55 \times 10^{-59}$

Çizelge 4.3'te kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,683 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,679 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama aşaması

Çizelge 4.4. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadrik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Cohen κ	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	Gwet AC ₂	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
	Scott π	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	Kripp. α	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	B-P	Zayıf	Zayıf	Zayıf	Orta	Ortalamadan iyiye	Orta
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Cohen κ	Önemsiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	Gwet AC ₂	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Scott π	Önemsiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	Kripp. α	Önemsiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	B-P	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
FLEISS'İNE GÖRE	Cohen κ	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf - Önemsiz	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf-Önemsiz
	Gwet AC ₂	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Scott π	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf - Önemsiz	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf-Önemsiz
	Kripp. α	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf - Önemsiz	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf-Önemsiz
	B-P	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf - Önemsiz	Zayıf-Önemsiz-Yetersiz	Zayıf	Zayıf-Önemsiz
ALTMAN'A GÖRE	Cohen κ	Önemsiz-Yetersiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	Gwet AC ₂	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Scott π	Önemsiz-Yetersiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	Kripp. α	Önemsiz-Yetersiz	Zayıf	Önemsiz	Önemsiz-Yetersiz	Zayıf	Önemsiz
	B-P	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman ölçeklerinde B-P katsayısının basit sıralı ağırlıkla hesaplandığında bir derece azaldığı, buna karşın Fleiss ölçeğinde bir seviye arttığı gözlenmiştir.

Katsayılar arasında en yüksek uyum değerine sahip Gwet AC₂, literatürde her iki ağırlık tipine göre hesaplandığında elde edilen sonuç bakımından Landis&Koch kıyaslama ölçeğine

göre “Önemli”, Fleiss kıyaslama ölçeğine göre “Ortalamadan iyiye” ve Altman kıyaslama ölçeğine göre “İyi” bir uyumu temsil etmektedir. Bu etiketleme standart hata bilgisinden yoksun bir etiketleme olduğundan kıyaslama Gwet Yöntemine göre de yapılmış ve incelenmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, katsayıların temsil ettiği uyum seviyelerinde herhangi bir fark gözlenmemiş olup, yalnızca Altman ölçeğinde, Landis&Koch ölçeğindeki ilk iki aralığı birleştirildiği için, uyum, hem “Önemsiz” hem de “Yetersiz” olarak değerlendirilmiştir.

Gwet’in kıyaslama metodu Landis&Koch ölçeğindeki aralıklara göre yapılandırıldığında, kuadratik ağırlıklar için yine Landis&Koch ve Altman ölçeği kullanıldığında, B-P katsayısının temsil ettiği uyum seviyesinde herhangi bir değişimin olmadığı, buna karşın diğer katsayıların bir seviye azaldığı gözlenmiştir. Landis&Koch ölçeğine göre yapılandırılan aralıklarla Fleiss ölçeği kullanıldığında ise, tüm katsayılar literatürdeki mevcut uyum seviyelerini korudukları gözlenmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında, kümülatif aitlik olasılıklarına göre yapılandırılmış Landis&Koch ölçeğinde ise kuadratik ağırlıklara göre uyum seviyesinin iki seviye birden düştüğü gözlenmiştir. Örneğin kuadratik ağırlıklarda Cohen κ için Landis&Koch “Önemsiz” seviyesini veriyor iken, basit sıralı ağırlıklarda hem “Önemsiz” hem de “Yetersiz” şeklinde de yorum yapılabilmektedir. Fleiss ölçeğinde, B-P katsayısı bir seviye azalırken, diğer katsayıların sabit kaldığı görülmüştür. Altman ölçeğinde ise tüm katsayılar bir seviye azalmıştır.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, katsayıların temsil ettiği uyum seviyelerinde herhangi bir fark gözlenmemiştir.

Gwet'in kıyaslama metodu Fleiss ölçeğindeki aralıklara göre yapılandırıldığında, kuadratik ağırlıklar için Landis&Koch ölçeği kullanıldığında, tüm katsayıların uyum seviyelerinin azaldığı gözlenmiştir. Gwet AC₂ katsayısı bir seviye azalırken, diğer katsayılar 2 seviye birden azalmıştır. Fleiss ölçeği kullanıldığında, tüm katsayılar literatürdeki mevcut uyum seviyelerini korudukları gözlenmiştir. Altman ölçeği kullanıldığında ise, tüm katsayılarının uyum seviyelerinin bir seviye azaldığı gözlenmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında, kümülatif aitlik olasılıklarına göre yapılandırılmış Landis&Koch ölçeğinde Gwet AC₂'nin temsil ettiği uyum seviyesinde bir seviye azalma görülürken, Cohen κ , Scott'ın π ve Krippendorff'un α katsayılarının temsil ettiği uyumda 2 seviye ve ayrıca B-P katsayısında 3 seviye birden azalma görülmüştür. Kümülatif olasılıklar Fleiss ölçeği ile yapılandırılıp, yine Fleiss ölçeği kullanıldığında, B-P katsayısı bir derece azalırken, diğer katsayıların sabit olduğu görülmüştür. Altman ölçeği kullanıldığında ise, B-P katsayısı 2 seviye birden azalırken, diğer katsayıların bir seviye azaldığı gözlenmiştir.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, katsayıların temsil ettiği uyum seviyelerinde herhangi bir fak gözlenmemiştir.

Gwet'in kıyaslama metodu Altman ölçeğindeki aralıklara göre yapılandırıldığında, kuadratik ağırlıklar için Landis&Koch ölçeği kullanıldığında, B-P katsayısının temsil ettiği uyum seviyesinde herhangi bir değişiklik olmadığı, buna karşılık, Gwet AC₂ katsayısının bir seviye azalırken, diğer katsayıların ikişer seviye yani literatürde “Zayıf” bir uyumu temsil ederken, Altman ölçeğine göre yapılandırılan kümülatif olasılıklarla Gwet yönteminde “hem “Yetersiz” hem de “Önemsiz” olarak yorumlanabilecek bir uyumu temsil ettiği gözlenmiştir. Fleiss ölçeği kullanıldığında, tüm katsayılar literatürdeki mevcut uyum seviyelerini korudukları gözlenmiştir. Altman ölçeği kullanıldığında ise, B-P katsayısının temsil ettiği uyum seviyesinde herhangi bir değişiklik olmadığı, buna karşılık diğer katsayıların uyum seviyelerinin bir seviye azaldığı gözlenmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında, kümülatif aitlik olasılıklarına göre yapılandırılmış Landis&Koch ölçeği kullanıldığında, Gwet AC₂ ve B-P katsayılarının bir seviye azalırken, diğer katsayıların ikişer seviye yani literatürde “Zayıf” bir uyumu temsil ederken, Altman ölçeğine göre yapılandırılan kümülatif olasılıklarla Gwet yönteminde “hem “Yetersiz” hem de “Önemsiz” olarak yorumlanabilecek bir uyumu temsil ettiği gözlenmiştir. Fleiss ölçeği kullanıldığında, B-P katsayısı bir derece azalırken, diğer katsayıların sabit olduğu görülmüştür. Altman ölçeği kullanıldığında ise, tüm katsayıların bir seviye azaldığı gözlenmiştir.

4.1.2. Dört kategorili durum

Landis ve Koch(1977a) çalışmalarında, Westlung ve Kurland (1953)’de raporlanan çoklu skleroz verisinin kullanmışlardır. Bu veride araştırmacılar, hastalığın coğrafi dağılımındaki olası farkları araştırmak için hasta gruplarının karşılaştırılması ile ilgilenmişlerdir. Farklı bölgelerdeki tüm muayeneler toplandıktan sonra, hastanın öyküsü ve hastaya yapılan teşhis bilinmeden, her bir nörologtan hastaları aşağıdaki kategorilerden birine sınıflamaası istenmiştir:

1. Kesin çoklu skleroz
2. Muhtemel çoklu skleroz
3. Mümkün çoklu skleroz(odds 50;50)
4. Şüpheli veya kesinlikle çoklu skleroz değil

Teşhisçiler arasındaki uyumu değerlendirmek için Winnipeg nörologları, New Orleans hasta kayıtlarının her birini yeniden gözden geçirip sınıflamışlardır. Aynı şekilde New Orleans nörologları da Winnipeg hasta kayıtlarının her birini gözden geçirip yeniden sınıflandırmıştır. Burada, yalnızca Winnipeg hastalarının sınıflandırılması verisi ele alınacaktır. Araştırma sonucunda, teşhisler arasında belirgin farklar tespit edilmiştir. Araştırma verisi Çizelge 4.5’te verilmiştir.

Çizelge 4.5. Deneklerin değerlendirici ve kategori bazında dağılımı

		Winnipeg Nöroloğu				Toplam	Yüzde
New Orleans Nöroloğu		Kesin çoklu skleroz	Muhtemel çoklu skleroz	Mümkün çoklu skleroz	Şüpheli veya kesinlikle çoklu skleroz değil		
Kesin çoklu skleroz		38	5	0	1	44	[29,5%]
Muhtemel çoklu skleroz		33	11	3	0	47	[31,5%]
Mümkün çoklu skleroz		10	14	5	6	35	[23,5%]
Şüpheli veya kesinlikle çoklu skleroz değil		3	7	3	10	23	[15,4%]
Toplam		84	37	11	17	149	[100%]
Yüzde		[56,4%]	[24,8%]	[7,4%]	[11,4%]	[100%]	

Çizelge 4.5'teki dağılım yüzdelерinden A ve B değerlendiricilerinin skorlamaları arasında 4 kategoride de bir fark olduğu görülmektedir ancak yüzdelikler tek başına bir gösterge değildir. Bu nedenle, Çizelge 4.5'teki veriler sıralı ölçeklerde 2 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, Çizelge 4.7'de sonuçlar verilmiştir.

Çizelge 4.6. Çizelge 4.5'in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar

		Kuadratik Ağırlıklar				Basit Sıralı Ağırlıklar			
		1	2	3	4	1	2	3	4
1		1,000	0,889	0,556	0,000	1,000	0,833	0,500	0,000
2		0,889	1,000	0,889	0,556	0,833	1,000	0,833	0,500
3		0,556	0,889	1,000	0,889	0,500	0,833	1,000	0,833
4		0,000	0,556	0,889	1,000	0,000	0,500	0,833	1,000

1.Kesin çoklu skleroz

2.Muhtemel çoklu skleroz

3.Mümkün çoklu skleroz(odds 50;50)

4.Şüpheli veya kesinlikle çoklu skleroz değil

Çizelge 4.7. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

		Uyum Katsayıları	Hesaplanan Katsayı Değeri	Std Hata	95% Güven Aralığı	p-Değeri
KUA DRAT İK	Cohen κ		0,525	0,060	0,406 - 0,643	4,82x10 ⁻¹⁵
	Gwet AC2		0,622	0,055	0,513 - 0,731	1,32x10 ⁻²¹

Çizelge 4.7. (devam) Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

KUADRATİK	Scott π	0,497	0,069	0,361 - 0,633	$2,34 \times 10^{-11}$
	Krippendorff' α	0,499	0,068	0,363 - 0,634	$1,79 \times 10^{-11}$
	Brenann-Prediger B-P	0,549	0,058	0,434 - 0,664	$8,13 \times 10^{-17}$
	Gözlenen uyum p_a	0,875	0,016	0,843 - 0,907	$2,3 \times 10^{-99}$
BASİT SIRALI	Cohen κ	0,476	0,057	0,364 - 0,588	$3,55 \times 10^{-14}$
	Gwet AC2	0,572	0,054	0,467 - 0,678	$4,25 \times 10^{-20}$
	Scott π	0,447	0,065	0,318 - 0,576	$1,78 \times 10^{-10}$
	Krippendorff' α	0,449	0,065	0,321 - 0,577	$1,36 \times 10^{-10}$
	Brenann-Prediger B-P	0,502	0,054	0,395 - 0,610	$2,63 \times 10^{-16}$
	Gözlenen uyum p_a	0,845	0,017	0,811 - 0,878	$3,72 \times 10^{-94}$

Çizelge 4.7'te kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,622 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,572 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama aşaması

Çizelge 4.8. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadrik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Cohen κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	Scott π	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	B-P	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Cohen κ	Orta	Ortalamadan iyiye	Orta	Zayıf	Zayıf	Zayıf
	Gwet AC_2	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Scott π	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	Kripp. α	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	B-P	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
FLEISS'E GÖRE	Cohen κ	Orta-Zayıf	Ortalamadan iyiye	Orta-İyi	Yetersiz-Önemsiz-Zayıf	Ortalamadan iyiye	Önemsiz - Zayıf
	Gwet AC_2	Orta-Zayıf	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Scott π	Yetersiz-Önemsiz-Zayıf	Zayıf	Orta-Zayıf-Önemsiz	Yetersiz-Önemsiz-Zayıf	Zayıf	Önemsiz - Zayıf
	Kripp. α	Zayıf	Zayıf	Orta-Zayıf-Önemsiz	Yetersiz-Önemsiz-Zayıf	Zayıf	Önemsiz - Zayıf
	B-P	Orta-Zayıf	Ortalamadan iyiye	Orta	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
ALTMAN'A GÖRE	Cohen κ	Orta	Ortalamadan iyiye	Orta	Zayıf	Zayıf	Zayıf
	Gwet AC_2	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Scott π	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	Kripp. α	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf	Zayıf
	B-P	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, basit sıralı ağırlıklılarla yapılan hesaplamada uyumun gücünün Gwet AC_2 katsayısında kuadratik ağırlıklılarla yapılan hesaplamalara göre bir derece azaldığı gözlenmiştir. Bu etiketleme standart hata bilgisinden yoksun bir etiketleme olduğundan kıyaslama Gwet Yöntemine göre de yapılmış ve incelenmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Landis&Koch ölçeğindeki aralıklara göre yapılandırıldığında, her iki ağırlık bazında katsayılarda sadece Cohen κ katsayısının temsil ettiği uyum seviyesinde basit sıralı ağırlıklarla hesaplama yapıldığında bir seviye azalma görülmüştür. Diğer katsayılar ağırlık tipi değişse bile temsil edilen uyum seviyesi bakımından aynı kalmıştır.

Gwet metoduna göre, kuadratik ağırlıklar için Landis&Koch ve Altman ölçeklerinde, Gwet AC₂, Scott'ın π 'sinde ve Krippendorff'un α katsayısında bir derece düşük uyum seviyesi gösterdiği, Cohen'in κ 'sının ve B-P katsayısının değişmediği gözlenmiştir. Fleiss ölçeğinin temsil ettiği uyum seviyesi literatürdekine göre, Scott'ın π 'si ve Krippendorff'un α katsayısında bir derece düşerken diğer katsayılar için değişmemiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında ise, Landis&Koch aralıklarına göre yapılandırılan kümülatif olasılıklar ile hesaplanan aitlik aralıkları baz alınıp, literatürdeki mevcut üç ölçekle bu aralıklar kullanılarak yorum yapıldığında, Gwet AC₂, ve B-P katsayısının aynı kaldığı, diğer katsayıların ise uyum seviyesinin bir seviye düştüğü gözlenmiştir. Örneğin Cohen κ için Landis&Koch'da "Orta" derecede tanımlanan uyum, Gwet metoduna göre yorumlandığında "Zayıf" derecede uyumu temsil etmiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda Landis&Koch ölçeğine göre Cohen κ 'nın mevcut uyum seviyesini korumakla birlikte basit sıralı ağırlıklara göre hesaplandığında 3 seviye birden azaldığı, Gwet AC₂ katsayısının mevcut uyum seviyesini korumakla birlikte bir seviye arttığı, Scott'ın π katsayısının değişmediği, Krippendorff α katsayısının mevcut uyum seviyesini korumakla birlikte 2 seviye azaldığı ve B-P katsayısının bir seviye arttığı gözlenmiştir. Fleiss ölçeğine göre, ağırlık tipi değiştiğinde uyum seviyelerinde herhangi bir değişim gözlenmemiştir. Altman ölçeğinde ise, Cohen κ 'nın temsil ettiği uyum seviyesinin iki seviye azaldığı, Gwet AC₂'nin değişmediği, Scott'ın π ve Krippendorff'un α katsayılarının bir seviye azaldığı ve B-P katsayısının bir seviye arttığı gözlenmiştir.

Fleiss aralıklarına göre yapılandırılan Gwet yöntemine göre, Landis&Koch ölçeği kullanıldığında, Cohen κ ve B-P katsayılarının mevcut uyum seviyesini korumakla birlikte literatürdeki uyum seviyesine göre bir seviye azaldığı, Gwet AC₂'nin 2 seviye azaldığı, Scott π 'nin 3 seviye birden azaldığı gözlenmiştir. Fleiss ölçeğine göre, Scott π ve Krippendorff'un α katsayılarının bir seviye azaldığı ve diğer katsayıların literatüre göre temsil ettiği uyum seviyelerinin değişmediği gözlenmiştir. Altman ölçeğine göre, Cohen κ 'nın mevcut uyum seviyesini korumakla birlikte literatürdeki uyum seviyesine göre bir seviye arttığı, Gwet AC₂'nin mevcut uyum seviyesini korumakla birlikte bir seviye azaldığı, Scott π ve Krippendorff'un α katsayılarının mevcut uyum seviyesini korumakla birlikte 2 seviye azaldığı ve B-P katsayısının uyum seviyesinin değişmediği gözlenmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında ise, Fleiss aralıklarına göre yapılandırılan kümülatif olasılıklar ile hesaplanan aitlik aralıkları baz alınıp, literatürdeki Landis%Koch ve Altman ölçekleriyle bu aralıklar kullanılarak yorum yapıldığında, Gwet AC₂ ve B-P katsayısının mevcut uyum seviyesini korumakla birlikte bir seviye arttığı, diğer katsayıların ise 2 seviye birden azaldığı gözlenmiştir. Fleiss ölçeğine göre, , Scott π ve Krippendorff'un α katsayılarının bir derece azaldığı, buna karşılık diğer katsayıların değişmediği gözlenmiştir.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Her iki ağırlık tipine göre hesaplama yapıldığında Cohen κ katsayısının kuadratik ağırlıkta temsil ettiği uyum seviyesine göre basit sıralı ağırlıkta bir seviye düşük bir uyum temsil ettiği, diğer katsayılarda ise ağırlık değişirken herhangi bir değişimin olmadığı tespit edilmiştir.

Altman aralıklarına göre yapılandırılan Gwet yöntemine göre ölçekler değerlendirildiğinde, kuadratik ağırlıklar için Landis&Koch ve Altman ölçeğine göre literatürdeki uyum seviyelerine göre, Cohen κ ve B-P katsayısının temsil ettiği uyumun değişmediği, diğer katsayıların ise bir seviye düştüğü gözlenmiştir. Fleiss ölçeğine bakıldığında ise Scott'ın π ve Krippendorff'un α katsayısının bir seviye azaldığı görülürken, diğer katsayılarda herhangi bir değişim gözlenmemiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamalarda ise Gwet yöntemine göre her üç ölçek tipinde Cohen κ , Scott'ın π 'si ve Krippendorff'un α katsayısında bir seviye düşüş olduğu, diğer katsayılar da herhangi bir değişikliğin olmadığı tespit edilmiştir.

4.1.3. Beş kategorili durum

Landis ve Koch (1977b) çalışmalarında, Holmquist, McMahan ve Williams (1967) çalışmasında raporlanan veriyi kullanmışlardır. Rahim serviksde karsinoma in situ'nun klasifikasyonunun değişkenliğini araştırmak için 7 patoloğ ayrı ayrı 118 slaytı aşağıdaki kategorilerden birine sınıflaması istenmiştir:

1. Negatif
2. Atipik yassı hierplazi
3. Karsinoma in situ
4. Erken stromal invazyonla birlikte yassı karsinoma
5. İnvazif karsinoma

Araştırma ve kategori ile ilgili detaylar Holmquist, McMahan ve Williams(1967) verilmiş olup, analiz, A, B, C, D, E, F ve G olarak etiketlenen 7 patoloğdan sadece A ve B patoloğlarının teşhis verisi kullanılarak yapılacaktır. Araştırma ham verisi EK-5'te verilmiş olup, kontenjans tablosu halinde gösterimi Çizelge 4.19'da verilmiştir.

Çizelge 4.9. Rahim serviksde karsinoma in situ'nun klasifikasyon verisi

Değerlendirici A	Değerlendirici B					Toplam	Yüzde
	1	2	3	4	5		
1	22	2	2	0	0	26	[22%]
2	5	7	14	0	0	26	[22%]
3	0	2	36	0	0	38	[32,2%]
4	0	1	14	7	0	22	[18,6%]
5	0	0	3	0	3	6	[5,1%]
Toplam	27	12	69	7	3	118	[100%]
Yüzde	[22,9%]	[10,2%]	[58,5%]	[5,9%]	[2,5%]	[100%]	

Çizelge 4.9'daki dağılım yüzdelerinden A ve B değerlendiricilerinin skorlamaları arasında bir fark olduğu görülmektedir. Bu nedenle, Çizelge 4.9'daki veriler sıralı ölçeklerde 2

değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, sırasıyla Çizelge 4.10'da sonuçlar verilmiştir.

Çizelge 4.10. Çizelge 4.9'in analizi için kullanılan kuadratik ve basit sıralı ağırlıklar

Kuadratik Ağırlıklar						Basit Sıralı Ağırlıklar					
	1	2	3	4	5		1	2	3	4	5
1	1,000	0,938	0,750	0,438	0,000		1,000	0,900	0,700	0,400	0,000
2	0,938	1,000	0,938	0,750	0,438		0,900	1,000	0,900	0,700	0,400
3	0,750	0,938	1,000	0,938	0,750		0,700	0,900	1,000	0,900	0,700
4	0,438	0,750	0,938	1,000	0,938		0,400	0,700	0,900	1,000	0,900
5	0,000	0,438	0,750	0,938	1,000		0,000	0,400	0,700	0,900	1,000

- 1.Negatif
2.Atipik yassı hierplazi
3.Karsinoma in situ
4.Erken stromal invazyonla birlikte yassı karsinoma
5.İnvazif karsinoma

Çizelge 4.11. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

	Uyum Katsayıları	Hesaplanan Katsayı Değeri	Std Hata	95% Güven Aralığı	p-Değeri
KUADRATİK	Cohen κ	0,779	0,041	0,698 - 0,860	1,289x10 ⁻³⁷
	Gwet AC2	0,906	0,017	0,871 - 0,940	3,442x10 ⁻⁸³
	Scott π	0,778	0,041	0,697 - 0,860	1,649x10 ⁻³⁷
	Krippendorff α	0,779	0,041	0,698 - 0,860	1,017x10 ⁻³⁷
	Brenann-Prediger B-P	0,871	0,021	0,828 - 0,913	4,955x10 ⁻⁷¹
	Gözlenen uyum p_a	0,968	0,005	0,957 - 0,978	3,256x10 ⁻¹⁴⁵
BASİT SIRALI	Cohen κ	0,735	0,044	0,649 - 0,821	4,77 x10 ⁻³⁶
	Gwet AC2	0,874	0,021	0,833 - 0,914	8,36 x10 ⁻⁷⁶
	Scott π	0,733	0,044	0,646 - 0,821	1,86 x10 ⁻³⁵
	Krippendorff α	0,735	0,044	0,647 - 0,822	1,16 x10 ⁻³⁵
	Brenann-Prediger B-P	0,834	0,024	0,785 - 0,882	1,44 x10 ⁻⁶⁵
	Gözlenen uyum p_a	0,953	0,007	0,940 - 0,967	8,3 x10 ⁻¹³⁴

Çizelge 4.11'de kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri<0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir.

Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,906 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,874 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama Aşaması

Çizelge 4.12. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
LİTERATÜRE GÖRE		Kuadratik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis& Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
	Cohen κ	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Kripp. α	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
LANDIS&KOCH' A GÖRE		GWET YÖNTEMİ					
		Landis& Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
	Cohen κ	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Kripp. α	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi

Çizelge 4.12. (devam)Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
FLEISS'E GÖRE	Cohen κ	Zayıf-Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi
	Scott π	Zayıf-Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Kripp. α	Zayıf-Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi
ALTMAN'A GÖRE	Cohen κ	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Kripp. α	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman ölçeklerine göre herhangi bir değişikliğin olmadığı ancak Fleiss ölçeğine göre basit sıralı ağırlıklılarla yapılan hesaplamada uyumun gücünün Cohen κ , Scott'ın π ve Krippendorff'un α katsayısında kuadratik ağırlıklılarla yapılan hesaplamalara göre bir derece azaldığı gözlenmiştir. Bu etiketleme standart hata bilgisinden yoksun bir etiketleme olduğundan kıyaslama Gwet Yöntemine göre de yapılmış ve incelenmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Landis&Koch ölçeğindeki aralıklara göre yapılandırıldığında, her iki ağırlık bazında katsayılar Landis&Koch ve Altman ölçeklerine göre, sadece B-P katsayısının temsil ettiği uyum seviyesinde basit sıralı ağırlıklılarla hesaplama yapıldığında bir seviye azalma görülmüştür. Diğer katsayılar ağırlık tipi değişse bile temsil edilen uyum seviyesi bakımından aynı kalmıştır.

Gwet metoduna göre, kuadratik ağırlıklar için her üç ölçek tipinde de, uyum katsayılarının literatürdeki ölçeklere göre temsil ettikleri uyumlarda herhangi bir değişikliğin olmadığı tespit edilmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında ise, Landis&Koch aralıklarına göre yapılandırılan kümülatif olasılıklar ile hesaplanan aitlik aralıkları baz alınıp, literatürdeki mevcut üç ölçekle bu aralıklar kullanılarak yorum yapıldığında, Landis&Koch ve Altman ölçekleri için sadece B-P katsayısının temsil ettiği uyum seviyesinde bir seviye azalma görülmüş olup, diğer katsayıların temsil ettikleri uyum seviyesi bakımından literatürden farklı olmadıkları görülmüştür. Fleiss ölçeği için, Cohen κ , Scott π ve Krippendorff α katsayıları bir seviye yükselirken, Gwet AC₂ sabit kalmış ve B-P katsayısının bir seviye azaldığı gözlenmiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda Landis&Koch kıyaslama ölçeği için Gwet AC₂ ve B-P katsayılarının basit sıralı ağırlıklarla hesaplandığında kuadratik ağırlıklara göre mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı, diğer katsayıların ise mevcut uyum seviyelerini korumakla birlikte bir seviye arttığı görülmüştür. Fleiss kıyaslama ölçeği için herhangi bir değişiklik gözlenmemiştir. Altman ölçeğine göre bakıldığında, Gwet AC₂ ve B-P katsayılarının basit sıralı ağırlıklarla hesaplandığında kuadratik ağırlıklara göre mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı, diğer katsayılarda ise ağırlık değiştiğinde temsil ettikleri uyum seviyeleri bakımından herhangi bir değişiklik gözlenmemiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında Fleiss ölçeğinin aralıklarıyla yapılandırılmış Gwet yönteminde Landis&Koch ölçeğinde Gwet AC₂ ve B-P katsayılarının literatürde temsil ettiği aynı uyum seviyesini koruduğu ancak diğer uyum katsayılarının mevcut uyum seviyelerini korumakla birlikte 2 seviye azaldığı, Altman ve Fleiss ölçeğinde ise benzer şekilde Gwet AC₂ ve B-P katsayılarının literatürde temsil ettiği aynı uyum seviyesini koruduğu ancak diğer uyum katsayılarının mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı görülmüştür.

Basit sıralı ağırlıklarla hesaplama yapıldığında ise, Fleiss aralıklarına göre, yine Fleiss ölçeği kullanılarak değerlendirme yapıldığında, uyum katsayılarının literatürde temsil ettikleri uyumlarla herhangi bir değişimin olmadığı tespit edilmiştir. Fleiss aralıklarına göre yapılandırılan kümülatif olasılıklarla Landis&Koch ve Altman ölçekleri kullanıldığında ise, tüm katsayıların literatürde temsil ettikleri uyum seviyelerinin bir seviye azaldığı görülmüştür.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Her iki ağırlık tipine göre hesaplama yapıldığında Landis&Koch ve Altman ölçeklerine göre sadece B-P katsayısının basit sıralı ağırlıkla hesaplandığında bir seviye azaldığı, diğer katsayıların ise sabit kaldığı gözlenmiştir. Fleiss ölçeğine bakıldığında ise Gwet AC₂'nin sabit kaldığı ancak diğer katsayıların temsil ettikleri uyum dereceleri bakımından bir seviye azaldığı gözlenmiştir.

Altman aralıklarına göre yapılandırılan Gwet yöntemine göre ölçekler değerlendirildiğinde, kuadratik ağırlıklar için mevcut üç ölçek için literatürde temsil edilen uyum seviyelerinin aynen korunduğu tespit edilmiştir.

Basit sıralı ağırlıklar kullanılarak hesap yapıldığında ise, Landis&Koch ve Altman ölçeğine göre, B-P katsayısının temsil ettiği uyumun bir seviye azaldığı görülmüştür. Fleiss ölçeğine göre ise, Gwet AC₂'nin sabit kaldığı, B-P katsayısının mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı ve diğer katsayıların da mevcut uyum seviyelerini korumakla birlikte bir seviye arttığı gözlenmiştir.

4.1.4. Altı kategorili durum

Swieten ve diğerleri(1987), Utrecht Üniversite Hastanesinden ve Rotterdam Dijkzigt Üniversite Hastanesinden sırasıyla tesadüfi olarak seçilen 50 hekim çiftiyle gerçeğe en yakın çok merkezli klinik deney durumunu simule etmek istemişlerdir. İki farklı hastaneden tesadüfi olarak seçilen 1000 doktor çifti, 1986 yılının Mart 1 ve Eylül 1 tarihleri arasında, 6 ay boyunca, hem nöroloji departmanındaki hem de sevk edilen doktor tarafından beyin enfarktüsü teşhisi konulan tüm hastaları kapsamaktadır. Her hastanede 50 adet olmak üzere

100 hasta değerlendirilmiş olup, nörolojik eksikliği yalnızca 24 saatten fazla sürenler çalışmaya dahil edilmiştir.

Engelliliğin kaydedilebilmesi için aşağıdaki 6 kategori oluşturulmuş ve doktorlara verilen bir eğitim ile açıklanmıştır.

0. Hiçbir semptom yok

1. Semptomlara rağmen önemli bir engellilik yok: tüm sıradan aktiviteleri ve görevleri gerçekleştirebilir
2. Hafif engellilik: daha önce gerçekleştirdiği tüm aktiviteleri gerçekleştiremiyor ancak yardım almadan kendi kendine bakabilir
3. Orta engellilik: biraz yardıma ihtiyacı var ancak yardım almadan yürüyebilir
4. Kısmen şiddetli engellilik: yardım almadan yürüyemez ve yardım almadan bedensel ihtiyaçlarını karşılayamaz
5. Şiddetli engellilik: yatalak, idrarını tutamaz, sürekli hemşire bakımı ve dikkatine ihtiyacı var.

Çalışmanın verileri kontenjans tablosu halinde gösterimi Çizelge 4.43'te verilmiştir.

Çizelge 4.13. Beyin enfarktüsü teşhisi verisi

Gözlemci 2	Gözlemci 1						Toplam	Yüzde
	0	1	2	3	4	5		
0	5	0	0	0	0	0	5	[5%]
1	0	6	2	0	0	0	8	[8%]
2	1	4	13	5	2	0	25	[25%]
3	0	0	6	9	4	0	19	[19%]
4	0	0	0	2	8	1	11	[11%]
5	0	0	0	0	8	24	32	[32%]
Toplam	6	10	21	16	22	25	100	[100%]
Yüzde	[6%]	[10%]	[21%]	[16%]	[22%]	[25%]	[100%]	

Çizelge 4.13'teki dağılım yüzdelerinden A ve B değerlendiricilerinin skorlamaları arasında bir fark olduğu görülmektedir. Bu nedenle, Çizelge 4.13'teki veriler sıralı ölçeklerde 2 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, sırasıyla Çizelge 4.15'te sonuçlar verilmiştir.

Çizelge 4.14. Çizelge 4.13'ün analizi için kullanılan kuadratik ve basit sıralı ağırlıklar

Kuadratik Ağırlıklar							Basit Sıralı Ağırlıklar					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1	0,96	0,84	0,64	0,36	0	1	0,93	0,8	0,6	0,33	0
1	0,96	1	0,96	0,84	0,64	0,36	0,93	1	0,93	0,8	0,6	0,33
2	0,84	0,96	1	0,96	0,84	0,64	0,8	0,93	1	0,93	0,8	0,6
3	0,64	0,84	0,96	1	0,96	0,84	0,6	0,8	0,93	1	0,93	0,8
4	0,36	0,64	0,84	0,96	1	0,96	0,33	0,6	0,8	0,93	1	0,93
5	0	0,36	0,64	0,84	0,96	1	0	0,33	0,6	0,8	0,93	1

0. Hiçbir semptom yok

1. Semptomlara rağmen önemli bir engellilik yok: tüm sıradan aktiviteleri ve görevleri gerçekleştirebilir

2. Hafif engellilik: daha önce gerçekleştirdiği tüm aktiviteleri gerçekleştiremiyor ancak yardım almadan kendi kendine bakabilir

3. Orta engellilik: biraz yardıma ihtiyacı var ancak yardım almadan yürüyebilir

4. Kısmen şiddetli engellilik: yardım almadan yürüyemez ve yardım almadan bedensel ihtiyaçlarını karşılayamaz

5. Şiddetli engellilik: yataklak, idrarını tutamaz, sürekli hemşire bakımı ve dikkatine ihtiyacı var.

Çizelge 4.15. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

	Uyum Katsayıları	Hesaplanan Katsayı Değeri	Std Hata	95% Güven Aralığı	p-Değeri
KUADRATİK	Cohen κ	0,907	0,020	0,868 - 0,946	$2,42 \times 10^{-68}$
	Gwet AC2	0,934	0,012	0,910 - 0,959	$1,87 \times 10^{-89}$
	Scott π	0,907	0,020	0,868 - 0,946	$2,46 \times 10^{-68}$
	Krippendorff α	0,907	0,020	0,868 - 0,947	$1,46 \times 10^{-68}$
	Brenann-Prediger B-P	0,925	0,013	0,898 - 0,951	$1,11 \times 10^{-85}$
	Gözlenen uyum p_a	0,982	0,003	0,976 - 0,989	$1,9 \times 10^{-150}$
BASİT SIRALI	Cohen κ	0,873	0,024	0,825 - 0,921	$1,53 \times 10^{-58}$
	Gwet AC2	0,906	0,016	0,875 - 0,937	$4,57 \times 10^{-78}$
	Scott π	0,873	0,024	0,825 - 0,921	$1,81 \times 10^{-58}$
	Krippendorff α	0,874	0,024	0,825 - 0,922	$1,07 \times 10^{-58}$
	Brenann-Prediger B-P	0,895	0,017	0,861 - 0,928	$6,46 \times 10^{-75}$
	Gözlenen uyum p_a	0,973	0,004	0,964 - 0,981	$7,1 \times 10^{-136}$

Çizelge 4.15'te kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,934 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,906 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama aşaması

Çizelge 4.16. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadratik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
FLEISS'E GÖRE	Cohen κ	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	Gwet AC ₂	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	Scott π	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi

Çizelge 4.16. (devam)Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

	Kripp. α	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	B-P	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
ALTMAN'A GÖRE	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, herhangi bir katsayıda değişim gözlenmemiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Landis&Koch ölçeğindeki aralıklara göre yapılandırıldığında, hem ağırlık hem de literatürdeki uyum seviyeleri bakımından katsayıların temsil ettiği uyum seviyeleri bakımından herhangi bir değişiklik gözlenmemiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda sadece Altman kıyaslama ölçeğinde tüm katsayıların temsil ettikleri uyum seviyesi bakımından bir seviye azaldığı tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise Landis&Koch ölçeği kullanıldığında, hem kuadratik hem de basit sıralı ağırlıklar için hesaplanan uyum katsayılarının uyum seviyelerinin bir seviye azaldığı görülmüştür. Fleiss ölçeklerinde literatürdeki uyum seviyeleri bakımından herhangi bir fark yokken, Altman ölçeğinde tüm katsayılar için mevcut uyum seviyelerini korumakla birlikte bir derece azalma görülmüştür.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Altman ölçeğindeki aralıklara göre yapılandırıldığında, hem ağırlık hem de literatürdeki uyum seviyeleri bakımından katsayıların temsil ettiği uyum seviyeleri bakımından herhangi bir değişiklik gözlenmemiştir.

4.1.5. Yedi kategorili durum

Atiya ve diğerleri (2003), Kadın Sağlık Çalışmalarında tutulan medikal kayıt incelemelerine dayanarak felç sınıflandırması ve engellilik durumundaki değerlendiriciler arası güvenilirliği değerlendirmişlerdir. 7 yıl boyunca 271 kaza meydana gelmiş olup, bunların 133 tanesi yeniden sınıflanmıştır. Çalışmanın, 7 sıralı kategoriden oluşan engellilik derecesi verisi kullanılacaktır. Analiz verisi Çizelge 4.18'de gösterilmiştir.

Çizelge 4.17. Engellilik derecesi sınıflandırması ham verisi

Birinci Sınıflama	İkinci Sınıflama							Toplam	Yüzde
	Hiç	Önemli Değil	Hafif	Orta	Ortadan şiddetliye	Şiddetli	Ölüm		
Hiç	4	6	0	0	0	0	0	10	[7,5%]
Önemli Değil	7	33	7	1	1	0	0	49	[36,8%]
Hafif	0	9	12	2	1	0	0	24	[18%]
Orta	0	5	9	6	3	0	0	23	[17,3%]
Ortadan şiddetliye	0	0	1	3	7	1	0	12	[9%]
Şiddetli	0	0	0	0	0	2	0	2	[1,5%]
Ölüm	0	0	0	0	0	0	13	13	[9,8%]
Toplam	11	53	29	12	12	3	13	133	[100%]
Yüzde	[8,3%]	[39,8%]	[21,8%]	[9%]	[9%]	[2,3%]	[9,8%]	[100%]	

Çizelge 4.17'deki dağılım yüzdelerinden 1. ve 2. sınıflama skorlamaları arasında çok da fazla bir fark olmadığı anlaşılabılır ancak yüzdeler tek başına bir gösterge değildir. Bu nedenle, Çizelge 4.17'deki veriler sıralı ölçeklerde 2 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, Çizelge 4.19'da sonuçlar verilmiştir.

Çizelge 4.18. Çizelge 4.17'nin analizi için kullanılan kuadratik ve basit sıralı ağırlıklar

Kuadratik Ağırlıklar								Basit Sıralı Ağırlıklar							
	1	2	3	4	5	6	7		1	2	3	4	5	6	7
1	1,00	0,97	0,89	0,75	0,56	0,31	0,00	1,00	0,95	0,86	0,71	0,52	0,29	0,00	
2	0,97	1,00	0,97	0,89	0,75	0,56	0,31	0,95	1,00	0,95	0,86	0,71	0,52	0,29	
3	0,89	0,97	1,00	0,97	0,89	0,75	0,56	0,86	0,95	1,00	0,95	0,86	0,71	0,52	
4	0,75	0,89	0,97	1,00	0,97	0,89	0,75	0,71	0,86	0,95	1,00	0,95	0,86	0,71	
5	0,56	0,75	0,89	0,97	1,00	0,97	0,89	0,52	0,71	0,86	0,95	1,00	0,95	0,86	
6	0,31	0,56	0,75	0,89	0,97	1,00	0,97	0,29	0,52	0,71	0,86	0,95	1,00	0,95	
7	0,00	0,31	0,56	0,75	0,89	0,97	1,00	0,00	0,29	0,52	0,71	0,86	0,95	1,00	

1. Hiç
2. Önemli Değil
3. Hafif
4. Orta
5. Ortadan şiddetliye
6. Şiddetli
7. Ölüm

Çizelge 4.19. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

	Uyum Katsayıları	Hesaplanan Katsayı Değeri	Std Hata	95% Güven Aralığı	p-Değeri
KUADRATİK	Cohen κ	0,885	0,025	0,836 - 0,934	$2,26 \times 10^{-69}$
	Gwet AC2	0,939	0,011	0,918 - 0,959	$5,96 \times 10^{-120}$
	Scott π	0,885	0,025	0,836 - 0,934	$2,83 \times 10^{-69}$
	Krippendorff' α	0,885	0,025	0,836 - 0,935	$1,70 \times 10^{-69}$
	Brenann-Prediger B-P	0,917	0,013	0,891 - 0,943	$9,32 \times 10^{-106}$
	Gözlenen uyum p_a	0,982	0,003	0,976 - 0,987	$3,94 \times 10^{-195}$
BASIT SIRALI	Cohen κ	0,847	0,029	0,789 - 0,905	$4,55 \times 10^{-59}$
	Gwet AC2	0,914	0,013	0,889 - 0,939	$1,35 \times 10^{-107}$
	Scott π	0,847	0,029	0,789 - 0,905	$6,28 \times 10^{-59}$
	Krippendorff' α	0,847	0,029	0,790 - 0,905	$3,78 \times 10^{-59}$
	Brenann-Prediger B-P	0,887	0,015	0,857 - 0,918	$1,43 \times 10^{-95}$
	Gözlenen uyum p_a	0,972	0,004	0,964 - 0,981	$2,02 \times 10^{-180}$

Çizelge 4.19'da kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,939 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,914 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama aşaması

Çizelge 4.20. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
LİTERATÜRE GÖRE		Kuadratik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
FLEISS'E GÖRE	Cohen κ	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	Gwet AC ₂	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	Scott π	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi

Çizelge 4.20. (devam)Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

	Kripp. α	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	B-P	Önemli-Neredeyse Mükemmel	Mükemmel	Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
ALTMAN'A GÖRE	Cohen κ	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Scott π	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Kripp. α	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, herhangi bir katsayıda değişim gözlenmemiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Landis&Koch ölçeğindeki aralıklara göre yapılandırıldığında, hem ağırlık hem de literatürdeki uyum seviyeleri bakımından katsayıların temsil ettiği uyum seviyeleri bakımından herhangi bir değişiklik gözlenmemiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda her üç ölçekte de tüm katsayıların temsil ettikleri uyum seviyesi bakımından herhangi bir değişiklik tespit edilememiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ölçeği için tüm katsayıların mevcut uyum seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Basit ağırlıklarla hesaplama yapıldığında ise, sadece Altman ölçeğinde tüm katsayıların mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı görülmüştür.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Gwet'in kıyaslama metodu Altman ölçeğindeki aralıklara göre yapılandırıldığında, hem ağırlık hem de literatürdeki uyum seviyeleri bakımından katsayıların temsil ettiği uyum seviyeleri bakımından herhangi bir değişiklik gözlenmemiştir.

4.2. Üç ve Daha Fazla Değerlendiricili Durumlar

Bu bölümde tamamı 5 sıralı kategoriden oluşan sırasıyla 3, 4, 5 ve 7 değerlendiricili örneklerin uyum katsayısı analizi yapılmıştır.

Uyum katsayısı analizi yapılan tüm çalışmalar sıralı 5 kategoriden oluştuğundan sırasıyla hepsinde Çizelge 4.10'da hesaplanan kuadratik ve basit sıralı ağırlıkların aynıları kullanılmıştır.

4.2.1. Üç değerlendirici

Alt bölüm 4.1.3'te ele alınan Landis ve Koch (1977b) çalışmasındaki A, B, C, D, E, F ve G olarak etiketlenen 7 patolojiden sadece A, E ve G patolojilerinin teşhis verisi kullanılarak yapılacaktır. Araştırma ham verisi EK-5'te verilmiş olup, kontenjans tablosu halinde gösterimi Çizelge 4.71'de verilmiştir.

Çizelge 4.21. Rahim serviksde karsinoma in situ'nun klasifikasyon verisi

Değerlendiriciler	1	2	3	4	5	Toplam
A	26	26	38	22	6	118
E	16	32	52	14	4	118
G	32	20	61	3	2	118
Ortalama	24,7	26	50,3	13	4	118,00

Çizelge 4.21'deki veriler sıralı ölçeklerde 3 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, sırasıyla Çizelge 4.22'de sonuçlar verilmiştir.

Çizelge 4.22. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

Yöntem	Katsayı	StdHata	95% G.A.	p-Değeri
Conger κ	0,766	0,031	0,704 - 0,828	$5,925 \times 10^{-48}$
Gwet AC ₂	0,903	0,013	0,878 - 0,928	$2,874 \times 10^{-98}$
Fleiss κ	0,764	0,032	0,701 - 0,827	$4,510 \times 10^{-47}$
Krippendorff α	0,765	0,027	0,712 - 0,818	$1,972 \times 10^{-54}$
Brenann-Prediger B-P	0,871	0,014	0,844 - 0,899	$3,528 \times 10^{-91}$
Gözlenen uyum p_a	0,968	0,004	0,961 - 0,975	$3,137 \times 10^{-166}$
Conger κ	0,714	0,033	0,648 - 0,780	$2,74 \times 10^{-42}$
Gwet AC ₂	0,867	0,016	0,836 - 0,898	$2,93 \times 10^{-86}$
Fleiss κ	0,712	0,034	0,644 - 0,779	$2,25 \times 10^{-41}$
Krippendorff α	0,713	0,029	0,655 - 0,770	$4,99 \times 10^{-48}$
Brenann-Prediger B-P	0,830	0,017	0,797 - 0,862	$6,59 \times 10^{-81}$
Gözlenen uyum p_a	0,952	0,005	0,943 - 0,961	$1,59 \times 10^{-151}$

Çizelge 4.22’de kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0’a çok yakın olduğu tespit edilmiştir. p-değeri<0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0’dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,903 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂’nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,867 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için de Gwet AC₂’dir.

Kıyaslama aşaması

Çizelge 4.23. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadrik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Conger κ	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Fleiss κ	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Kripp. α	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Conger κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Fleiss κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Kripp. α	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
FLEISS'E GÖRE	Conger κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Gwet AC_2	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
	Fleiss κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Kripp. α	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	B-P	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok İyi
ALTMAN'A GÖRE	Conger κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Fleiss κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Kripp. α	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	B-P	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman'da herhangi bir katsayıda değişim gözlenmemişken, Fleiss ölçeğinde Conger κ , Fleiss κ ve Krippendorff α katsayılarında bir seviye azalış tespit edilmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Conger κ , Fleiss κ ve Krippendorff α katsayılarında basit sıralı ağırlıklarda Fleiss ölçeğine göre bir seviye azalış tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için katsayıların temsil ettiği uyum seviyeleri literatürdeki ile değişiklik göstermezken, Fleiss ölçeğinde, Conger κ , Fleiss κ ve Krippendorff α katsayılarında bir seviye azalış tespit edilmiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, sadece mevcut üç ölçekte de tüm katsayıların mevcut uyum seviyeleri ile literatürdekiler arasında herhangi bir fark gözlenmemiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda her üç ölçekte de tüm katsayıların temsil ettikleri uyum seviyesi bakımından herhangi bir değişiklik tespit edilememiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için tüm katsayıların mevcut uyum seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, Gwet AC₂ ve B-P katsayılarının literatürdeki uyum seviyelerini korurken, Conger κ , Fleiss κ ve Krippendorff α katsayılarında bir seviye azalış tespit edilmiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, sadece Fleiss ölçeğinde tüm katsayıların mevcut uyum seviyelerini korudukları gözlemlenirken, Landis&Koch ve Altman ölçeklerinde tüm katsayıların mevcut uyum seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Conger κ , Fleiss κ ve Krippendorff α katsayılarında basit sıralı ağırlıklarda Fleiss ölçeğine göre bir seviye azalış tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için katsayıların temsil ettiği uyum seviyeleri literatürdeki ile değişiklik göstermezken, Fleiss ölçeğinde, Conger κ , Fleiss κ ve Krippendorff α katsayılarında bir seviye azalış tespit edilmiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, sadece mevcut üç ölçekte de tüm katsayıların mevcut uyum seviyeleri ile literatürdekiler arasında herhangi bir fark gözlenmemiştir.

4.2.2. Dört değerlendirici

4 değerlendirici ve 5 kategorili durumdaki yum katsayılarını analiz etmek için Gwet (2014: 371) Sticklefish verisi kullanılmıştır. Araştırma ham verisi EK-6'da verilmiştir.

Çizelge 4.24. Sticklefish verisi

Değerlendiriciler	1	2	3	4	5	Toplam
Değerlendirici 1	10	0	11	1	7	29
Değerlendirici 2	10	2	11	1	5	29
Değerlendirici 3	10	1	9	3	6	29
Değerlendirici 4	12	0	6	3	8	29
Ortalama	10,5	0,8	9,3	2	6,5	29,00

Çizelge 4.24'teki veriler sıralı ölçeklerde 3 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup Çizelge 4.25'te sonuçlar verilmiştir.

Çizelge 4.25. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

Yöntem	Katsayı	StdHata	95% G.A.	p-Değeri
Conger κ	0,734	0,067	0,597 - 0,871	$1,16 \times 10^{-14}$
Gwet AC ₂	0,762	0,040	0,679 - 0,844	$1,75 \times 10^{-20}$
Fleiss κ	0,734	0,067	0,597 - 0,871	$1,21 \times 10^{-14}$

Çizelge 4.25. (devam) Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

Krippendorff α	0,736	0,055	0,624 - 0,848	$9,11 \times 10^{-17}$
Brenann-Prediger B-P	0,682	0,054	0,572 - 0,793	$4,62 \times 10^{-16}$
Gözlenen uyum p_a	0,921	0,014	0,893 - 0,948	$1,22 \times 10^{-35}$
Conger κ	0,697	0,071	0,551 - 0,842	$1,44 \times 10^{-13}$
Gwet AC ₂	0,727	0,044	0,636 - 0,817	$6,49 \times 10^{-19}$
Fleiss κ	0,696	0,071	0,55 - 0,842	$1,56 \times 10^{-13}$
Krippendorff α	0,699	0,059	0,579 - 0,819	$1,77 \times 10^{-15}$
Brenann-Prediger B-P	0,649	0,057	0,533 - 0,765	$4,64 \times 10^{-15}$
Gözlenen uyum p_a	0,902	0,016	0,869 - 0,934	$1,90 \times 10^{-33}$

Çizelge 4.22’de kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0’a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0’dan farklı olduğu söylenebilir.

Uyumun sıfırdan farklı olduğu tespit edildikten sonra, miktarı ile ilgili uyum katsayılarının nicelikleri incelenmiştir. Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,762 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂’nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,727 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂’nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂’dir.

Kıyaslama aşaması

Çizelge 4.26. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadrik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Conger κ	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC_2	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Fleiss κ	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
	Kripp. α	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
	B-P	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Conger κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	Fleiss κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
	B-P	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
FLEISS'E GÖRE	Conger κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Gwet AC_2	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Fleiss κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Kripp. α	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	B-P	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
ALTMAN'A GÖRE	Conger κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Mükemmel	İyi
	Fleiss κ	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Mükemmel	İyi
	B-P	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman'da herhangi bir katsayıda değişim gözlenmemişken, Fleiss ölçeğinde Gwet AC_2 'nin bir seviye azaldığı tespit edilmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Conger κ , Fleiss κ katsayılarının basit sıralı ağırlıklarda Landis&Koch ve Altman ölçeğine göre bir seviye azaldığı, diğerlerinin ise ağırlık değişiminden etkilenmediği tespit edilmiştir. Fleiss ölçeğine göre ise, sadece B-P katsayısının ağırlık farkından etkilenmediği, geri kalan uyum katsayılarının bir seviye azaldığı tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için yalnızca bir seviye azalmış olup, diğer katsayıların temsil ettiği uyum seviyeleri literatürdeki ile değişiklik gözlenmemiştir. Fleiss ölçeğinde, Conger κ , Fleiss κ ve Krippendorff α katsayılarında katsayılarında bir seviye artış tespit edilirken, B-P katsayısının sabit kaldığı ve Gwet AC₂ katsayısının bir seviye azaldığı gözlenmiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, Gwet AC₂ ve Krippendorff α katsayısında literatürdeki uyum seviyeleri bakımından herhangi bir değişim gözlenmezken, diğer katsayılarda bir seviye azalış gözlenmiştir. Fleiss ölçeğine göre ise literatürdeki uyum seviyelerinde herhangi bir değişim olmamıştır.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda her üç ölçekte de tüm katsayıların temsil ettikleri uyum seviyesi bakımından herhangi bir değişiklik tespit edilememiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için tüm katsayıların mevcut uyum seviyelerini seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, diğer katsayılar mevcut uyum seviyelerini korurken, yalnızca Gwet AC₂'nin bir seviye azaldığı görülmüştür.

Basit sıralı ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için tüm katsayıların mevcut uyum seviyelerini seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, tüm katsayıların literatürde temsil ettikleri uyumu koruduğu tespit edilmiştir.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Conger κ , Fleiss κ ve B-P katsayılarında basit sıralı ağırlıklarda Landis&Koch ve Altman ölçeğine göre bir seviye azalış tespit edilmiştir. Fleiss ölçeğine göre Conger κ , Fleiss κ ve B-P katsayılarında basit sıralı ağırlıklarda bir seviye azalış tespit edilirken, Gwet AC₂ ve Krippendorff α katsayılarında bir seviye artış tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için katsayıların temsil ettiği uyum seviyeleri için sadece B-P katsayısında bir seviye azalan uyum gözlenirken, diğer katsayıların temsil ettikleri uyum seviyesinin literatürdekine göre değişiklik göstermediği tespit edilmiştir. Fleiss ölçeğinde, Gwet AC₂ katsayısının uyum seviyesi bir seviye azalırken, diğer katsayılarda değişiklik gözlenmemiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, Landis&Koch ve Altman ölçeği için Gwet AC₂ ve Krippendorff α katsayıları değişmezken, diğer katsayıların bir seviye azaldığı gözlenmiştir. Fleiss ölçeğine göre ise, Gwet AC₂ ve Krippendorff α katsayılarının bir seviye arttığı ve diğer katsayıların değişmediği gözlenmiştir.

4.2.3. Beş değerlendirici

Alt bölüm 4.1.3'te ele alınan Landis ve Koch (1977b) çalışmasındaki A, B, C, D, E, F ve G olarak etiketlenen 7 patolojiden sadece A, D, E, F ve G patolojilerinin teşhis verisi kullanılarak yapılacaktır. Araştırma ham verisi EK-5'te verilmiş olup, kontenjans tablosu halinde gösterimi Çizelge 4.92'de verilmiştir.

Çizelge 4.27. Rahim serviksde karsinoma in situnun klasifikasyon verisi

Değerlendiriciler	1	2	3	4	5	Toplam
A	26	26	38	22	6	118
D	38	48	23	8	1	118
E	16	32	52	14	4	118
F	63	30	20	1	4	118
G	32	20	61	3	2	118
Ortalama	35	31,2	38,8	9,6	3,4	118,00

Çizelge 4.27'deki veriler sıralı ölçeklerde 5 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, Çizelge 4.29'da sonuçlar verilmiştir.

Çizelge 4.28. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

Yöntem	Katsayı	StdHata	95% G.A.	p-Değeri
Conger κ	0,626	0,045	0,537 - 0,715	$1,04 \times 10^{-29}$
Gwet AC ₂	0,830	0,018	0,793 - 0,867	$1,29 \times 10^{-78}$
Fleiss κ	0,616	0,048	0,522 - 0,711	$2,80 \times 10^{-27}$
Krippendorff α	0,617	0,040	0,539 - 0,695	$2,46 \times 10^{-33}$
Brenann-Prediger B-P	0,783	0,021	0,741 - 0,825	$1,61 \times 10^{-69}$
Gözlenen uyum p_a	0,946	0,005	0,935 - 0,956	$1,37 \times 10^{-147}$
Conger κ	0,579	0,043	0,493 - 0,664	$1,82 \times 10^{-28}$
Gwet AC ₂	0,786	0,020	0,747 - 0,826	$1,35 \times 10^{-72}$
Fleiss κ	0,568	0,046	0,478 - 0,659	$4,90 \times 10^{-26}$
Krippendorff α	0,569	0,038	0,493 - 0,645	$8,11 \times 10^{-32}$
Brenann-Prediger B-P	0,736	0,022	0,693 - 0,779	$5,65 \times 10^{-65}$
Gözlenen uyum p_a	0,926	0,006	0,914 - 0,938	$5,65 \times 10^{-139}$

Çizelge 4.28'de kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0'a çok yakın olduğu tespit edilmiştir. p-değeri < 0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0'dan farklı olduğu söylenebilir.

Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,830 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en

düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,786 değeriyle Gwet AC₂'dir. Gwet AC₂'nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂'nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂'dir.

Kıyaslama aşaması

Çizelge 4.29. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

Kıyaslama Ölçeği							
		Kuadrik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Conger κ	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Mükemmel	İyi
	Fleiss κ	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	B-P	Önemli	Mükemmel	İyi	Önemli	Ortalamadan iyiye	İyi
GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Conger κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Fleiss κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	B-P	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
FLEISS'E GÖRE	Conger κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Gwet AC ₂	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi
	Fleiss κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Kripp. α	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	B-P	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi

Çizelge 4.29. (devam)Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

ALTMAN'A GÖRE	Conger κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Fleiss κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Orta	Ortalamadan iyiye	Orta	Orta	Mükemmel	Orta
	B-P	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman'da B-P katsayısında değişim gözlenmezken, diğer katsayıların basit sıralı ağırlıklarda bir seviye azaldığı gözlenmiştir. Fleiss ölçeğine göre ise, B-P katsayısı basit sıralı ağırlıklarda bir seviye azalırken, diğer katsayıların değişmediği gözlenmiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Gwet AC_2 'nin her üç ölçekte de bir derece azaldığı gözlenmiştir. Diğer katsayılarda üç ölçek için de ağırlık değişiminden etkilenmediği tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için Gwet AC_2 ve B-P katsayısının değişmediği, diğer katsayıların bir seviye azaldığı gözlenmiştir. Fleiss ölçeğine göre, yalnızca B-P katsayısında mevcut seviyesini korumakla birlikte bir seviye azaldığı ve diğer katsayıların literatürdeki ölçeklere göre uyum seviyelerinin değişmediği gözlenmiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda her üç ölçekte de tüm katsayıların temsil ettikleri uyum seviyesi bakımından herhangi bir değişiklik tespit edilememiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için tüm katsayıların mevcut uyum seviyelerini seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, diğer katsayılar mevcut uyum seviyelerini korurken, yalnızca Gwet AC₂'nin bir seviye azaldığı görülmüştür.

Basit sıralı ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için tüm katsayıların mevcut uyum seviyelerini seviyelerinin korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, tüm katsayıların literatürde temsil ettikleri uyumu koruduğu tespit edilmiştir.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu temsil edilen uyum seviyelerinde herhangi bir farklılık tespit edilememiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçeği için katsayıların temsil ettiği uyum seviyelerinde Gwet AC₂ ve B-P katsayılarının uyum seviyelerinin aynı kaldığı, diğer katsayıların ise temsil ettikleri uyum seviyelerinin bir seviye azaldığı gözlenmiştir. Fleiss ölçeğinde B-P katsayısı mevcut uyum seviyesini korumakla birlikte bir seviye azalırken, diğer katsayıların temsil ettiği uyum seviyeleri aynı kalmıştır.

Basit ağırlıklarla hesaplama yapıldığında ise, Landis&Koch ve Altman ölçeği için Gwet AC₂'nin temsil ettiği uyum seviyesinde bir derece artış gözlenirken, diğer katsayılar sabit kalmıştır. Fleiss ölçeğine göre ise, Krippendorff α katsayısının bir derece arttığı ve B-P

katsayılarının mevcut uyum seviyesini korurken bir seviye arttığı gözlenmiş, buna karşılık diğer katsayıların değişmediği gözlenmiştir.

4.2.4. Yedi değerlendirici

Alt bölüm 4.1.3'te ele alınan Landis ve Koch (1977b) çalışmasındaki A, B, C, D, E, F ve G olarak etiketlenen 7 patoloğun tümünün teşhis verisi kullanılarak yapılacaktır. Araştırma ham verisi EK-5'te verilmiştir.

Çizelge 4.30. Rahim serviksde karsinoma in situnun klasifikasyon verisi

Değerlendiriciler	1	2	3	4	5	Toplam
A	26	26	38	22	6	118
B	27	12	69	7	3	118
C	31	42	37	6	2	118
D	38	48	23	8	1	118
E	16	32	52	14	4	118
F	63	30	20	1	4	118
G	32	20	61	3	2	118
Ortalama	33,3	30	42,9	8,7	3,1	118,00

Çizelge 4.30'daki veriler sıralı ölçeklerde 7 değerlendirici için uygun olan uyum katsayılarıyla analiz edilmiş olup, sırasıyla Çizelge 4.31'de sonuçlar verilmiştir.

Çizelge 4.31. Kuadratik ve basit sıralı ağırlıklara göre hesaplanmış ağırlıklandırılmış uyum katsayıları, standart hata, %95 güven aralığı ve p-değerleri

Yöntem	Katsayı	StdHata	95% G.A.	p-Değeri
Conger κ	0,646	0,040	0,568 - 0,724	$6,44 \times 10^{-32}$
Gwet AC ₂	0,851	0,016	0,820 - 0,882	$2,87 \times 10^{-85}$
Fleiss κ	0,641	0,041	0,559 - 0,722	$2,22 \times 10^{-30}$
Krippendorff α	0,641	0,034	0,575 - 0,708	$1,00 \times 10^{-37}$
Brenann-Prediger B-P	0,805	0,018	0,770 - 0,840	$2,58 \times 10^{-76}$
Gözlenen uyum p_a	0,951	0,004	0,943 - 0,960	$6,45 \times 10^{-154}$
Conger κ	0,601	0,039	0,525 - 0,677	$2,26 \times 10^{-30}$
Gwet AC ₂	0,811	0,017	0,778 - 0,844	$1,46 \times 10^{-79}$
Fleiss κ	0,595	0,040	0,516 - 0,674	$8,29 \times 10^{-29}$
Krippendorff α	0,596	0,033	0,531 - 0,660	$4,36 \times 10^{-36}$
Brenann-Prediger B-P	0,761	0,018	0,725 - 0,797	$4,60 \times 10^{-72}$
Gözlenen uyum p_a	0,933	0,005	0,923 - 0,943	$1,53 \times 10^{-145}$

Çizelge 4.31’de kuadratik ve basit sıralı ağırlıklara göre iki değerlendirici durumunda hesaplanan uyum katsayıları ve gözlenen uyum, bu katsayılarla ilişkin standart hata, güven aralığı ve p-değerleri mevcuttur. P-değerlerinin tamamının 0’a çok yakın olduğu tespit edilmiştir. p-değeri<0,025 olduğundan H_0 hipotezini reddedilir ve H_1 hipotezi kabul edilir. Yani, bu çalışmada iki değerlendirici arasındaki uyumu temsil eden tüm kestirimlerin 0’dan farklı olduğu söylenebilir.

Buna göre, kuadratik ağırlıklara göre yapılan hesaplamada, en yüksek uyum katsayısı 0,851 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂’nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Basit sıralı ağırlıklara göre yapılan hesaplamada ise en yüksek uyum katsayısı 0,811 değeriyle Gwet AC₂’dir. Gwet AC₂’nin standart hatası da yine bu 5 katsayısının içinde en düşük standart hatadır. Dolayısıyla Gwet AC₂’nin güven aralığı en dar olan ve yığın parametresini diğer 4 katsayıya göre daha iyi bir temsilini sağlayan bir uyum katsayısı olduğu tespit edilmiştir.

Bu çalışma için uyumu temsil eden katsayılardan en güvenilir her iki ağırlık tipi için Gwet AC₂’dir.

Kıyaslama aşaması

Çizelge 4.32. Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

		Kıyaslama Ölçeği					
		Kuadratik Ağırlıklar			Basit Sıralı Ağırlıklar		
		Landis&Koch	Fleiss	Altman	Landis&Koch	Fleiss	Altman
LİTERATÜRE GÖRE	Conger κ	Önemli	Ortalamadan iyiye	İyi	Önemli	Ortalamadan iyiye	İyi
	Gwet AC ₂	Neredeyse Mükemmel	Mükemmel	Çok İyi	Neredeyse Mükemmel	Mükemmel	Çok İyi
	Fleiss κ	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Önemli	Ortalamadan iyiye	İyi	Orta	Ortalamadan iyiye	Orta
	B-P	Önemli	Mükemmel	Çok İyi	Önemli	Mükemmel	İyi

Çizelge 4.32. (devam)Literatürdeki kıyaslama ölçekleri ve Gwet yöntemine göre uyum seviyeleri

GWET YÖNTEMİ							
LANDIS&KOCH'A GÖRE	Conger κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Fleiss κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	B-P	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
FLEISS'E GÖRE	Conger κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Gwet AC_2	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi
	Fleiss κ	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	Kripp. α	Orta-Önemli	Ortalamadan iyiye	Orta-İyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
	B-P	Önemli-Neredeyse Mükemmel	Mükemmel	İyi-Çok iyi	Orta-Önemli	Ortalamadan iyiye	Orta-İyi
ALTMAN'A GÖRE	Conger κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Gwet AC_2	Neredeyse Mükemmel	Mükemmel	Çok İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi
	Fleiss κ	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	Kripp. α	Orta	Ortalamadan iyiye	Orta	Orta	Ortalamadan iyiye	Orta
	B-P	Önemli	Ortalamadan iyiye-Mükemmel	İyi	Önemli	Ortalamadan iyiye-Mükemmel	İyi

Literatürdeki ölçekler

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, Landis&Koch ve Altman ölçeklerine göre, Fleiss κ ve Krippendorff α katsayılarının bir seviye azaldığı gözlenirken, diğer katsayılar da herhangi bir değişiklik olmamıştır. Fleiss ölçeğine göre ise, uyum seviyeleri bakımından ağırlık değiştiğinde herhangi bir değişiklik gözlenmemiştir.

Landis&Koch kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu, sadece Gwet AC_2 'nin her üç ölçekte de bir derece azaldığı gözlenmiştir. Diğer katsayılar da üç ölçek için de ağırlık değişiminden etkilenmediği tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ölçeği için, Gwet AC₂ ve B-P katsayılarının değişmediği, buna karşılık diğer katsayıların bir seviye azaldığı gözlenmiştir. Fleiss ölçeğine göre, B-P katsayısının mevcut uyum seviyesinin korumakla birlikte bir seviye azaldığı ancak diğer katsayıların literatürde temsil ettikleri uyum seviyelerini korudukları gözlenmiştir. Altman ölçeğine göre ise, sadece Gwet AC₂ katsayısının uyum seviyesinin değişmediği, buna karşılık diğer katsayıların bir seviye azaldığı gözlenmiştir.

Basit sıralı ağırlıklara göre hesaplama yapıldığında, Landis&Koch ve Altman ölçeğine göre Conger κ ve Gwet AC₂ katsayılarının literatüre göre temsil ettikleri uyum seviyesinin bir derece azaldığı, buna karşın diğer katsayıların uyum seviyelerinde herhangi bir değişim olmadığı tespit edilmiştir. Fleiss ölçeğine göre, Gwet AC₂ ve B-P katsayılarının mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı, buna karşılık diğer katsayıların sabit kaldığı tespit edilmiştir.

Fleiss kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Fleiss'e göre yapılandırılan Gwet yönteminde, her iki ağırlık tipi kullanılarak hesaplanan sonuçlarda her üç ölçekte de B-P katsayısı haricindeki tüm katsayılarla temsil ettikleri uyum seviyesi bakımından herhangi bir değişiklik tespit edilmezken, basit sıralı ağırlıklarla hesaplama yapıldığında, B-P katsayısının Landis&Koch ölçeğinde “Önemli” seviyesini korurken bir seviye azalarak “Orta” seviyede bir uyumu temsil ettiği görülmüştür. Benzer şekilde, Altman ölçeğinde de “İyi” seviyesini korurken, bir seviye azalarak “Orta” seviyede bir uyum temsil ettiği görülmüştür. Fleiss ölçeğinde ise bir seviye azaldığı tespit edilmiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ölçeğine göre, B-P katsayı haricindeki katsayılar mevcut uyum seviyelerini koruyup bir seviye azalırken, B-P katsayısı mevcut uyum seviyesini korumakla birlikte bir seviye artmıştır. Altman ölçeğine göre ise, B-P katsayı haricindeki katsayılar mevcut uyum seviyelerini koruyup bir seviye azalırken, B-P katsayısı mevcut uyum seviyesini korumakla birlikte bir seviye azalmıştır. Fleiss ölçeğinde ise, tüm katsayıların literatürde temsil ettikleri uyumu koruduğu tespit edilmiştir.

Basit sıralı ağırlıklarla hesaplama yapıldığında, Landis&Koch ve Altman ölçekleri için, Fleiss κ ve Krippendorff'un α katsayılarının mevcut uyum seviyelerini korumakla birlikte bir derece arttığı ancak diğer katsayıların mevcut uyum seviyelerini korumakla birlikte bir seviye azaldığı görülmüştür. Fleiss ölçeğinde ise, tüm katsayıların literatürde temsil ettikleri uyumu koruduğu tespit edilmiştir.

Altman kıyaslama yöntemine göre yapılandırılan Gwet yöntemi

Literatürdeki mevcut ölçeklerle kıyaslama yapıldığında, kuadratik ve basit sıralı ağırlıklılarla yapılan hesaplamalar sonucu Gwet AC_2 katsayısı Landis&Koch ve Altman ölçeklerine göre bir seviye azalırken, Fleiss ölçeğine göre mevcut uyum seviyesini korumakla birlikte bir seviye azalmıştır. Diğer katsayıların temsil edilen uyum seviyelerinde ağırlık değiştiğinde herhangi bir farklılık gözlenmemiştir.

Literatürdeki ölçeklerle uyum seviyeleri kıyaslandığında ise kuadratik ağırlıklarla hesaplama yapıldığında, Landis&Koch ölçeğine göre Gwet AC_2 ve B-P katsayılarının uyum seviyeleri değişmezken, diğer katsayıların temsil ettikleri uyum seviyelerinin bir derece azaldığı gözlenmiştir. Fleiss ölçeğine göre Gwet AC_2 ve B-P katsayıları mevcut uyum seviyesini korumakla birlikte bir seviye azalmışken, diğer katsayılar da herhangi bir değişim gözlenmemiştir. Altman ölçeğine göre ise, Gwet AC_2 'nin temsil ettiği uyum seviyesi literatüre göre değişmezken, diğer katsayıların temsil ettiği uyum seviyelerinin bir seviye azaldığı gözlenmiştir.

Basit ağırlıklarla hesaplama yapıldığında ise, Landis&Koch ve Altman ölçeği için Conger κ ve Gwet AC_2 'nin temsil ettiği uyum seviyesinde bir derece azalma gözlenirken, diğer katsayılar sabit kalmıştır. Fleiss ölçeğine göre ise, mevcut uyum seviyesini korumakla birlikte Gwet AC_2 ve B-P katsayılarının temsil ettikleri uyum seviyelerinin bir derece azaldığı gözlenmiş, diğer katsayıların literatüre göre temsil ettiği uyum seviyelerinde ise herhangi bir değişim gözlenmemiştir.

5. TARTIŞMA VE SONUÇ

Bu çalışmada, sıralı ölçekler içeren iki veya daha fazla değerlendiricinin katıldığı değerlendiriciler arası güvenirlik çalışmalarında, gerçekteki uyumu kestiren uyum katsayıları, ağırlıklandırılmış versiyonlarının varyansları ve eksik skorelama durumundaki varyansları ayrıca hesaplanan bu katsayıların karşılaştırma yöntemleri incelenmiş, söz konusu yöntemler temsil ettikleri uyumun gücü bakımından karşılaştırılmıştır. Gerçek hayattan alınan farklı denek, değerlendirici ve kategori sayılarına sahip veriler üzerinde farklı formlardaki uyum katsayıları hesaplanmış ve farklı sonuçlar elde edilmiştir. Bu farklılıklar, verilen farklı yapıda olmasına dayanmaktadır.

Genel olarak bir değerlendiriciler-arası güvenirlik çalışmasında uyum hesaplamasının avantaj ve dezavantajları aşağıdaki şekilde sıralanabilir:

- Bu çalışmada, deneklerin denek popülasyonu N'den tesadüfi olarak seçildiği ve değerlendiricilerin koşullu seçildiği varsayılmıştır. Bu genellikle gerçek hayatta oldukça sık karşılaşılabilecek bir durum olduğundan bu tezdeki formüller bu tarz durumlar için kullanılması bakımından avantajlıdır.
- Popüler istatistiksel yazılım programları (SPSS, SAS, MINITAB vb.) uyum katsayılarının ağırlıklandırılmış versiyonlarını içermemektedir. Ancak R yazılımında kullanılmak üzere bu kodlar mevcuttur ve ayrıca Agreestat 13.3 yazılımı sayesinde de kolaylıkla hesaplanabilmektedir.
- Kayıp verli skorelama durumunda varyans hesabının elle yapılması denek başında işlemler gerektirdiğinden oldukça zordur. Yazılım olmaksızın elle varyans hesabı yapılması, karmaşıklığı ve oldukça fazla sayıda işlem gerektirdiği için işlem hatası olasılığına sahiptir ve araştırmanın sonuçlarına olumsuz bir şekilde etkileyebilir.
- Ağırlık seçimi tamamen araştırmacının seçimine bırakılmıştır. Teorik olarak ağırlıklandırılmış uyum katsayılarının, ağırlıklandırılmamış versiyonlarına göre daha büyük çıkması beklenir ancak bu, kuadratik veya basit sıralı ağırlıklar kullanıldığında beklenemez.
- Uyumun gücünü niteleyen literatürdeki mevcut ölçekler ele alınmış olup, bu ölçeklerin anataj ve dezavantajları incelenmiştir. Bu ölçekler haricinde, Gwet alternatif kıyaslama yöntemi incelenmiş ve istatistiksel olarak daha sağlam olduğu görülmüştür.

Elde edilen bulgulardan çıkarılan sonuçlara ise aşağıdaki gibidir:

- Bu tezde kullanılan veriler gerçek hayattan alınan verilerle yapılmış çalışmalardan elde edilmiş olup, tezin 2. bölümünde yer alan uyum katsayılarının hesaplanması amacıyla kullanılmıştır.
- Gerçek veri kullanıldığında, hem 2 hem de 3 ve fazla değerlendiricili çalışmalarda en yüksek uyum değerini veren katsayının Gwet AC₂ katsayısı olduğu tespit edilmiştir. Ayrıca, bu katsayı tüm çalışmalarda en küçük varyansa ve en dar güven aralığına sahiptir.
- Brennan-Prediger'in B-P katsayısı (1981) formülündeki beklenen uyum değeri $1/q$ olduğundan, kategori sayısından direkt olarak etkilenmektedir. 2 değerlendiricili verilerde kategori sayısı arttıkça daha iyi bir uyumu temsil ettiği görülmüştür.
- Krippendorff α katsayısının 2 değerlendiricili durumlarda eksik veriyle hesaplanamamıştır. Fakat 3 ve daha fazla değerlendirici durumda eksik veriyle hesaplanabilmiştir.
- Kıyaslama yöntemleri ele alındığında genel olarak Gwet'in kümülatif aitlik olasılıklarını kullanan yöntemi literatürdeki ölçeklere göre bir seviye daha düşük uyum gösterdiği tespit edilmiştir ancak verinin doğasına göre bunun her zaman bu şekilde olmayacağı da tespit edilmiştir.
- Landis&Koch ve Altman kıyaslama ölçekleri genellikle uyumun gücünü benzer aralıklarla ifade ederken, Fleiss kıyaslama aralıklarına göre hesaplanan kümülatif aitlik olasılıkları değişik ölçeklerde oldukça geniş aralıkta uyum güçleri göstermektedir. Bu da uyumun yorumlanabilmesini oldukça zorlaştırmaktadır. Fleiss'in kıyaslama ölçeği, uyumun gücü, zayıf-ortalama-mükemmel uyum kategorilerinde değerlendirmek istendiğinde araştırmacılara önerilmiştir.
- Farklı kıyaslama ölçeklerine göre hesaplanan kümülatif aitlik olasılıkları genellikle hesaplanmasında baz alınan ölçek için aynı uyum seviyelerini göstermektedir (Örneğin, Gwet yöntemi için Landis&Koch kıyaslama ölçeği aralıkları kullanıldıysa, kümülatif aitlik olasılıklarının gösterdiği uyumun gücüyle, katsayı değerinin direk aralığa göre yorumlandığı literatürdeki Landis&Koch ölçeğinin gösterdiği uyum gücünün aynı çıktığı görülmüştür).
- Gwet'in standart hatayı da kapsayan istatistiksel kıyaslama modeli, değişik tasarımlara sahip değerlendiriciler arası güvenilirlik çalışmalarının karşılaştırılmasında kullanılabilir.

Bu yöntemin, karşılaştırma yaparken istatistiksel olarak çok daha sağlam bir değerlendirme sunduğu gözlenmiştir.

- Uyum katsayılarının güven aralıkları, herhangi bir ölçekle birlikte değerlendirildiği zaman, Gwet'in kümülatif aitik olasılıklarını kullanan yönteminin verdiği sonuçlarla benzer sonuçlar verdiği görülmüştür.
- Denek, değerlendirici ve kategori sayıları baz alındığında gerçekteki uyumu en iyi şekilde tespit eden bir kestirime sahip uyum katsayısı, simülasyon çalışması yapılmadığı için bu tezde önerilememiştir. Gelecek çalışmalarda, farklı denek, değerlendirici ve kategori kümeleri ele alınarak simülasyon çalışması yapılabilir ve uyum katsayılarının bu değişkenlerle hesaplandığında temsil ettiği uyum seviyesi gözlenebilir.

KAYNAKLAR

- Agresti, A. (1988), "A Model Agreement Between Ratings on an Ordinal Scale." *Biometrics*, 46, 293-302.
- Altman, D.G.(1991). *Practical Statistics for Medical Research*. Chapman and Hall
- Atiya, M., et al. "Interobserver agreement in the classification of stroke in the Women's Health Study." *Stroke; a journal of cerebral circulation* 34.2 (2003): 565-567.
- Bademci, V. (2011). "Kuder-Richardson 20, Cronbach'ın Alfası, Hoyt'un Varyans Analizi, Genellenirlik Kuramı Ve Ölçüm Güvenirliği Üzerine Bir Çalışma" *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 17,173-193.
- Benini, R. (1901). "Principii di Demongraphia: Manuali Barbera Di Scienze Giuridiche Sociali e Politiche" (No.29). *Firenze, Italy: G. Barbera*
- Bennett, E.M., Albert, R. and Goldstein, A.C. (1954), "Communications through limied response questioning." *Public Opinion Quaterly*, 18, 303-308.
- Berry, K.J. and Mielke, Jr., P. W. (1988), "A generalization of Cohen's Kappa Agreement Measure To Interval Measurement and Multiple Raters", *Educational and Psychological Measurement*, 48, 921-933.
- Bland, M.J., Altman, D.G.(1986), "Statistical Methods for Assesing Agreement Between two Methods of Clinical Measurement", *Lancet*, 1, 307-310
- Bland, M.J., Altman, D.G.(1996), "Statistics Notes: Measurement error.", *British Medical Journal*, 312, p.1654
- Brennan, R.L. and Pediger, D.J.(1981). "Coefficient Kappa: some uses, misuses, and alternatives." *Educational and Psychological Measurement*, 41, 687-699.
- Byrt, T., Bishop, J. and Carlin, J.B.(1993). "Bias, prevalence and Kappa." *Journal of Clinical Epidemiology*, 46, 423-429.
- Carletta, J.(1996). "Assessing Agreement on Classification Tasks: the Kappa Statistics." *Computational Linguistics*, 22, 1-6.
- Carmines, E.G. and Zeller, R. A. (1979), *Reliability and Validity Assessment*, Sage Publications.

- Cicchetti, D. V. and Feinstein, A. R. (1990). "High Agreement but low Kappa: II. Resolving the paradoxes.", *Journal of Clinical Epidemiology*, 43, 551-558.
- Cohen, J. (1960). "A Coefficient of agreement for nominal scales." *Educational and Psychological Measurement*, 20, 37-46.
- Cohen, J. (1968). "Weighted kappa: Nominal scale agreement with provision for scaled agreement of partial credit." *Psychological Bulletin*, 70, 213-220.
- Conger, A. J. (1980), "Integration and Generalization of Kappas for Multiple Raters" *Psychological Bulletin*, 88, 322-328.
- Everitt, B. S. (1992). *The Analysis of Contingency Tables (2nd Ed.)* Chapman and Hall, London.
- Feinstein, A. R. and Cicchetti, D. V. (1990), "High Agreement but low Kappa: I. The problems of two paradoxes.", *Journal of Clinical Epidemiology*, 43, 543-549.
- Fleiss, J. L. (1971). "Measuring nominal scale agreement among many raters", *Psychological Bulletin*, 76, 378-382.
- Fleiss, J. L. (1981). *Statistical Methods for Raters and Proportions*. John Wiley & Sons.
- Fleiss, J. L., Cohen, J. and Everitt, B. S. (1969). "Large sample standard errors of kappa and weighted kappa." *Psychological Bulletin*, 72, 323-327.
- Fleiss, J. L., Levin, B. and Paik, M. C. (2003). *Statistical Methods for Raters and Proportions (3rd. Ed.-Chapter 18)*. Wiley Series in Probability and Statistics.
- Fleiss, J. L., Nee, J. C. M. and Landis, J. R. (1979). "The large sample variance of kappa in the case of different sets of raters", *Psychological Bulletin*, 86, 974-977.
- Gwet, K. L. (2001). *Handbook of Inter-rater Reliability*, 1th Edition, (Gaithersburg: StatAxis Publishing)
- Gwet, K. L. (2001). *Statistical Tables for Inter-Rater Agreement*. (Gaithersburg: StatAxis Publishing)
- Gwet, K. L. (2008a). "Computing inter-rater reliability and its variance in the presence of high agreement." *British Journal of Mathematical and Statistical Psychology*, 61, 29-48.
- Gwet, K. L. (2008b). "Variance estimation of nominal-scale inter-rater reliability with random selection of raters.", *Psychometrika*, 73, 407-430.

- Gwet, K. L. (2010b). *The Practical Guide to Statistics: Basic Concepts, Methods, and Meaning*, Advanced Analytics, LLC.
- Gwet, K. L. (2014). *Handbook of Inter-rater Reliability*, 4th Edition, Advanced Analytics, LLC
- Guttman, L. (1945). "A basis for analyzing test-retest reliability". *Psychometrika*, 10, 255-282
- Holley, J. W. and Guilford, J. P. (1964). "A note on the G index of agreement." *Educational and Psychological Measurement*, 24, 749-753.
- Janson, S. and Vegelius, J. (1979). "On generalizations of the G index and the PHI coefficient to nominal scales." *Multivariate Behavioral Research*, 14, 255-269.
- Janson, H. and Olsson, U. (2001). "A Measure of Agreement for Interval or Nominal Multivariate Observations" *Educational and Psychological Measurement*, 61, 277-289.
- Janson, H. and Olsson, U. (2004). "A Measure of Agreement for Interval or Nominal Multivariate Observations by Different Sets of Judges" *Educational and Psychological Measurement*, 64, 62-70.
- Kılıç, S. (2009). "*Ölçümlerin Uyumluluğu Ve Tıptaki Uygulamaları*", Yüksek lisans Tezi, Çukurova Üniversitesi, Adana.
- Krippendorff, K. (1970). "Estimating the reliability, systematic error, and random error of interval data" " *Educational and Psychological Measurement*, 30, 61-70.
- Krippendorff, K. (1978). "Reliability of binary attribute data", *Biometrics*, 34, 142-144.
- Krippendorff, K. (2004). Reliability in Content Analysis: Some Common Misconceptions and Recommendations. *Human Communication Research*, 30 (3), 411-433. <http://dx.doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Landis, J. R., Koch G. (1977a). "The Measurement of Observer Agreement for Categorical Data", *Biometrics*, Vol. 33 No.1, 159-174
- Landis, J. R., Koch G. (1977b). "An Application of Hierarchical Kappa-Type Statistics in the Assessment of Majority Agreement among Multiple Observers", *Biometrics*, 33, No.2, 363-374
- Maxwell, A. E. (1977). "Coefficient of agreement between observers and their interpretation.", *British Journal of Psychiatry*, 130, 79-83.

Scott, W. A. (1955). "Reliability of content analysis: the case of nominal scale coding.", *Public Opinion Quarterly*, XIX, 321-325.

Swieten, J. C., et al. "Interobserver agreement for the assessment of handicap in stroke patients." *Stroke; a journal of cerebral circulation* 19.5 (1988): 604-607.

Walsh et al. (2014), Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. *PeerJ* 2:e651; DOI 10.7717/peerj.651

EKLER

EK-1. Kuadratik (Karesel) ve basit sıralı ağırlıkların hesaplanmasına ilişkin örnek bir uygulama

Bir değerlendiriciler-arası güvenilirlik deneyinde, r adet değerlendirici tarafından sıralı 5 kategoride (1. evre, 2. evre, 3. evre, 4. evre ve 5. evre) n adet deneğin skorlandığı düşünölsün. Bu örneğe ilişkin kuadratik ve basit sıralı ağırlıklar sırasıyla Çizelge 1.1 ve 1.2’de gösterilmiştir.

Kuadratik ağırlıkların hesaplanması

$w_{11}=1$	$w_{12}=1-(1-2)^2/(5-1)^2$ =0,9375	$w_{13}=1-(1-3)^2/(5-1)^2$ =0,75	$w_{14}=1-(1-4)^2/(5-1)^2$ =0,4375	$w_{15}=1-(1-5)^2/(5-1)^2=0$
$w_{21}=1-(2-1)^2/(5-1)^2$ =0,9375	$w_{22}=1$	$w_{23}=1-(2-3)^2/(5-1)^2$ =0,9375	$w_{24}=1-(2-4)^2/(5-1)^2$ =0,75	$w_{25}=1-(2-5)^2/(5-1)^2$ =0,4375
$w_{31}=1-(3-1)^2/(5-1)^2$ =0,75	$w_{32}=1-(3-2)^2/(5-1)^2$ =0,9375	$w_{33}=1$	$w_{34}=1-(3-4)^2/(5-1)^2$ =0,9375	$w_{35}=1-(3-5)^2/(5-1)^2$ =0,75
$w_{41}=1-(4-1)^2/(5-1)^2$ =0,4375	$w_{42}=1-(4-2)^2/(5-1)^2$ =0,75	$w_{43}=1-(4-3)^2/(5-1)^2$ =0,9375	$w_{44}=1$	$w_{45}=1-(4-5)^2/(5-1)^2$ =0,9375
$w_{51}=1-(5-1)^2/(5-1)^2$ =0	$w_{52}=1-(5-2)^2/(5-1)^2$ =0,4375	$w_{53}=1-(5-3)^2/(5-1)^2$ =0,75	$w_{54}=1-(5-4)^2/(5-1)^2$ =0,9375	$w_{55}=1$

Buradan kuadratik ağırlıklar Çizelge 2.1’deki gibi olur;

Çizelge 1.1: Sıralı 5 kategori için kuadratik ağırlıklar

	Evre 1	Evre 2	Evre 3	Evre 4	Evre 5
Evre 1	1	0,9375	0,75	0,4375	0
Evre 2	0,9375	1	0,9375	0,75	0,4375
Evre 3	0,75	0,9375	1	0,9375	0,75
Evre 4	0,4375	0,75	0,9375	1	0,9375
Evre 5	0	0,4375	0,75	0,9375	1

Basit sıralı ağırlıkların hesaplanması

Çizelge 1.2’deki ağırlıkları hesaplamak için Eş. 2.2 ve Eş. 2.3’teki ağırlık hesaplama formülleri kullanılmıştır. Buna göre öncelikli olarak Eş. 2.3’ten aşağıdaki M_{kl} değerleri bulunur.

EK-1. (devam)Kuadratik (Karesel) ve basit sıralı ağırlıkların hesaplanmasına ilişkin örnek bir uygulama

$$\begin{aligned}
 M_{12} &= \binom{2-1+1}{2} = 1 & M_{21} &= \binom{2-1+1}{2} = 1 & M_{31} &= \binom{3-1+1}{2} = 3 \\
 M_{13} &= \binom{3-1+1}{2} = 3 & M_{23} &= \binom{3-2+1}{2} = 1 & M_{32} &= \binom{3-2+1}{2} = 1 \\
 M_{14} &= \binom{4-1+1}{2} = 6 & M_{24} &= \binom{4-2+1}{2} = 3 & M_{34} &= \binom{4-3+1}{2} = 1 \\
 M_{15} &= \binom{5-1+1}{2} = 10 & M_{25} &= \binom{5-2+1}{2} = 6 & M_{35} &= \binom{5-3+1}{2} = 3 \\
 M_{41} &= \binom{4-1+1}{2} = 6 & M_{51} &= \binom{5-1+1}{2} = 10 \\
 M_{42} &= \binom{4-2+1}{2} = 3 & M_{52} &= \binom{5-2+1}{2} = 6 \\
 M_{43} &= \binom{4-3+1}{2} = 1 & M_{53} &= \binom{5-3+1}{2} = 3 \\
 M_{45} &= \binom{5-4+1}{2} = 1 & M_{54} &= \binom{5-4+1}{2} = 1
 \end{aligned}$$

M_{kl} değerleri hesaplandıktan sonra benzer şekilde Eş. 2.2'deki M_{1q} değeri Eş. 2.3'teki formüle göre $M_{1q} = \binom{5-1+1}{2} = 10$ olarak hesaplanır. Daha sonra, basit sıralı ağırlıklar Eş. 2.2'ye göre aşağıdaki gibi elde edilir:

$$\begin{aligned}
 w_{11} &= 1 & w_{12} &= 1-(1/10)=0,9 & w_{13} &= 1-(3/10)=0,7 & w_{14} &= 1-(6/10)=0,4 & w_{15} &= 1-(10/10)=0,0 \\
 w_{21} &= 1-(1/10)=0,9 & w_{22} &= 1 & w_{23} &= 1-(1/10)=0,9 & w_{24} &= 1-(3/10)=0,7 & w_{25} &= 1-(6/10)=0,4 \\
 w_{31} &= 1-(3/10)=0,7 & w_{32} &= 1-(1/10)=0,9 & w_{33} &= 1 & w_{34} &= 1-(1/10)=0,9 & w_{35} &= 1-(3/10)=0,7 \\
 w_{41} &= 1-(6/10)=0,4 & w_{42} &= 1-(3/10)=0,7 & w_{43} &= 1-(1/10)=0,9 & w_{44} &= 1 & w_{45} &= 1-(1/10)=0,9 \\
 w_{51} &= 1-(10/10)=0,0 & w_{52} &= 1-(6/10)=0,4 & w_{53} &= 1-(3/10)=0,7 & w_{54} &= 1-(1/10)=0,9 & w_{55} &= 1
 \end{aligned}$$

Elde edilen w_{kl} basit sıralı ağırlıkları Çizelge 2.2'de verilmiştir.

Çizelge 2.2: Basit sıralı ağırlıklar

	Evre 1	Evre 2	Evre 3	Evre 4	Evre 5
Evre 1	1,0	0,9	0,7	0,4	0,0
Evre 2	0,9	1,0	0,9	0,7	0,4
Evre 3	0,7	0,9	1,0	0,9	0,7
Evre 4	0,4	0,7	0,9	1,0	0,9
Evre 5	0,0	0,4	0,7	0,9	1,0

Evre 1 ile Evre 2 veya Evre 2 ile Evre 1 gibi birbirine çok yakın iki kategori için sınıflama yapıldığında, uyumsuzluğun 0,9 gibi önemli bir ağırlığa sahip olduğu görülebilir. Benzer

EK-1. (devam)Kuadratik (Karesel) ve basit sıralı ağırlıkların hesaplanmasına ilişkin örnek bir uygulama

şekilde, Evre 1 ve Evre 4 gibi uzak iki kategori için sınıflama yapıldığında da bu uyuşmazlığın, 0,9'dan nicelik olarak daha küçük bir ağırlık olan 0,4 ağırlığına sahip olduğu görülmektedir. Kısmi uyum kavramı gereği, birbirine yakın olan kategorilere yapılan sınıflamalardaki uyuşmazlıklara daha yüksek ağırlık atandığı bu örnek yoluyla gösterilmiştir.

EK-2. Kappa: karesi alınmış öklid uzaklıklarının bir fonksiyonu

Ağırlıklandırılmış kappa katsayısı, 1960 yılında Cohen tarafından nominal ölçekler için geliştirilen uyum katsayısının 1968 yılında yine Cohen tarafından genişletilmesiyle elde edilmiş ve sıralı ölçekler için de kullanılabilir hale gelmiştir. Bu katsayı karesi alınmış Öklid uzaklıklarının bir fonksiyonudur.

A ve B değerlendiricilerinin n tane deneği 1 ve 2 gibi iki kategoriden birine sınıflamak zorunda olduğu bir durum düşünölsün. Bu kategorilerin birine bir deneği sınıflamak, bu deneğe iki elemanlı bir vektör atamakla eşdeğerdir. A_{ik} , eğer değerlendirici A i. deneği k. kategoriye atamışsa 1, diğör hallerde 0 değerini alan bir ikili değışken olsun. A değerlendiricisi için i. deneğin kategorizasyonu, i ye atanan bir (A_{i1}, A_{i2}) vektör skoruyla eş anlamlıdır. Bu vektör skoru A_i olarak etiketlenecektir. Benzer şekilde, B_i vektörü de $B_i = (B_{i1}, B_{i2})$ şeklinde B değerlendiricisiyle ilişkili olarak tanımlanabilir. O zaman, Kappa katsayısı aşağıdaki gibi sunulabilir:

$$\hat{K}_C = \frac{\frac{1}{n} \sum_{i=1}^n d^2(A_i, B_i)}{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d^2(A_i, B_j)} \quad (2.1)$$

burada $d^2(A_i, B_j)$, A_i 'den B_j 'ye karesi alınmış Öklid Uzaklığını temsil etmektedir. Eş. 2.4 kategori sayısı $q > 2$ olsa bile aynı kalır. Q kategorili durumda A_i ve B_j q-boyutlu vektörler haline gelir. Örneğin, A değerlendiricisi ve i. denekle ilişkilendirilen vektör skoru:

$$A_i = (A_{i1}, \dots, A_{i2}, \dots, A_{ik}) \quad (2.2)$$

olarak verilsin.

Eş. 2.4 sınıflama tipindeki skorlar yerine, değerlendiriciler x_1, x_2 ve x_3 gibi 3 aralıklı skarlardan birine atama yaptığında bile kullanılabilir. Bu durumda A_i vektörü tek bir meydana geliş için 1 ve diğör iki meydana geliş için 0 içeren 3 elemanlı bir vektör olmayacaktır. Onun yerine, A_i , (ya x_1 ya x_2 ya da x_3 : i. deneğe hangisinin atandığına bağılı olarak) tek bir değıer alacaktır. Daha önce de göröldüğü gibi, bu katsayı x_1, x_2, x_3 gibi basit kategori etiketler olursa, Cohen'in (1960) klasik Kappa katsayısıyla tıpa tıp aynıdır.

EK-2. (devam)Kappa: karesi alınmış öklid uzaklıklarının bir fonksiyonu

q aralıklı veriyle birlikte kullanıldığında $(x_1, \dots, x_k, \dots, x_q)$, Eş. 2.3'ü izleyen Kappa katsayısına neden olur:

$$\hat{\kappa}'_C = 1 - \frac{\sum_{k,l}^q p_{kl}(x_k - x_l)^2}{\sum_{k,l}^q p_{k+}p_{+l}(x_k - x_l)^2} \quad (2.3)$$

Burada,

$$p_{kl} = n_{kl}/n \quad (2.4)$$

p_{kl} : A ve B değerlendiricilerin sırasıyla x_k ve x_l skorlarına atadıkları deneklerin oranı,

$$p_{k+} = n_{k+}/n \quad (2.5)$$

p_{k+} ; A değerlendiricisinin x_k skorunu atadığı deneklerin oranı,

$$p_{+l} = n_{+l}/n \quad (2.6)$$

p_{+l} ; B değerlendiricisinin x_l skorunu atadığı deneklerin oranıdır.

Bu sonuç Eş 2.4'ten A değerlendiricisinin i . deneğe atadığı tek aralıklı x_l ($l = 1, 2, \dots, q$) skoruyla Eş 2.4'teki A_i vektör skorunun yer değiştirmesiyle elde edilir.

Aralıklı veri ve hiçbir eksik skorlaması olmayan tam bir veri seti ile ilgilenildiğinde, Kappa katsayısını hesaplamak için Eş. 2.3 kullanılabilir. Ancak, pratikte veri setleri, bazı değerlendiricilerin deneklerin kısıtlı sayısı üzerinde skorlama üretmesi ile birlikte sıklıkla tam değildir. Bu nedenle, skorlama yapan değerlendiricilerin uygun sayıdasıyla ilgili olarak p_{kl} , p_{k+} ve p_{+l} gibi farklı denek oranlarını garantileyen Kappa'nın daha genel bir formülasyonu Gwet (2014;80) tarafından önerilmiştir. Örneğin, A ve B değerlendiricilerin sırasıyla k ve l kategorilerine sınıflandırdığı deneklerin oranı, her iki değerlendiricinin skorlamış olduğu deneklerle ilgili olarak değerlendirilmelidir. Eğer bir denek, 2 değerlendiriciden yalnızca biri tarafından skorlandıysa, o zaman bu oranın hesaplanmasından hariç tutulacaktır. Bu denekleri hariç tutmamak, Kappa katsayısının olduğundan az gösterilmesini neden olacaktır.

EK-3. İki değerlendiricili durum için örnek uygulama örneği

Burada iki değerlendirici arasındaki uyumu ölçmek amacıyla geliştirilmiş uyum katsayılarının hesaplanmasıyla ilgili bir uygulama verilmiştir. Uygulamada iki değerlendirici 11 birimi A, B ve C gibi sıralı 3 kategoride değerlendirmiş ve veriler Çizelge 3.1’de verilmiştir. Çizelge 3.1’in skorumaya verisini ele alınsın. Bu veri seti için ağırlıklandırılmış ve ağırlıklandırılmamış uyum katsayıları ayrı ayrı hesaplanmıştır.

Çizelge 3.1. 11 deneğin 2 değerlendirici tarafından skorumaları

Birimler	Değerlendirici A	Değerlendirici B
1	A	
2	B	C
3	C	C
4	C	C
5	B	B
6	B	
7	A	A
8	A	B
9	B	B
10	B	B
11		C

Kaynak: Gwet [2014:80]

Ham veriler Çizelge 3.2’deki gibi bir kontenjans tablosu haline getirilir. p_{kl} değerlerinin hesaplamasında kullanılacak olan n_{kl}, n_{k+}, n_{+l} değerleri Çizelge 3.2’de verilmiştir.

Çizelge 3.2. n_{kl}, n_{k+}, n_{+l} değerleri

Değerlendirici 1	Değerlendirici 2			Kayıp Skorlar	n_{k+}
	A	B	C		
A	$n_{11} = 1$	$n_{12} = 1$	$n_{13} = 0$	1	3
B	$n_{21} = 0$	$n_{22} = 3$	$n_{23} = 1$	1	5
C	$n_{31} = 0$	$n_{32} = 0$	$n_{33} = 2$	0	2
Kayıp Skorlar	0	0	1	0	1
n_{+l}	1	4	4	2	11

Çizelge 3.2’de verilen n_{kl}, n_{k+}, n_{+l} değerleri kullanılarak Eş. 2.4, 2.7 ve 2.10’a göre p_{kl} değerleri hesaplanmış ve Çizelge 3.3’te verilmiştir.

EK-3. (devam)İki değerlendiricili durum için örnek uygulama örneği

Çizelge 3.3. p_{kl} değerleri

Değerlendirici 1	Değerlendirici 2			Kayıp Skorlar	p_{k+}
	A	B	C		
A	0,091	0,091	0,000	0,091	0,273
B	0,000	0,273	0,091	0,091	0,455
C	0,000	0,000	0,182	0,000	0,182
Kayıp Skorlar	0,000	0,000	0,091	0,000	0,091
p_{+l}	0,091	0,364	0,364	0,182	1,000

Eş. 2.8'e göre $p_{AB} = 8/11$, Eş. 2.11'e göre $p_A = 10/11$ ve Eş. 2.14'e göre $p_B = 10/11$ olarak hesaplanmıştır.

Çizelge 3.3'de verilen p_{kl}, p_{k+}, p_{+l} değerleri kullanılarak Eş. 2.7, 2.10 ve 2.13'e göre $p'_{kl}, p'_{k+}, p'_{+l}$ nispi oranları hesaplanmış ve ayrıca, Eş. 2.56 ve 2.57'ye göre π_k ve π_l değerleri de hesaplanarak Çizelge 3.4'te verilmiştir.

Çizelge 3.4. p'_{kl} değerleri

Değerlendirici 1	Değerlendirici 2			p'_{k+}	π_k
	A	B	C		
A	0,125	0,125	0,000	0,300	0,206
B	0,000	0,375	0,125	0,500	0,472
C	0,000	0,000	0,250	0,200	0,322
p'_{+l}	0,111	0,444	0,444		
π_l	0,206	0,472	0,322		

Cohen'in κ katsayısını hesaplamak için gereken $p'_{k+} \times p'_{+l}$ çarpımı Çizelge 3.5'te, Scott'ın π katsayısını hesaplamak için gereken ve Eş. 2.38 ve 2.39'ya göre hesaplanan π'_k ve π'_l değerlerinin $\pi'_k \times \pi'_l$ çarpımı da Çizelge 3.6 verilmiştir.

Çizelge 3.5. $p'_{k+} \times p'_{+l}$ değerleri

	A	B	C
A	0,033	0,133	0,133
B	0,056	0,222	0,222
C	0,022	0,089	0,089

EK-3. (devam)İki değerdendiricili durum için örnek uygulama örneđi

Çizelge 3.6. $\pi'_k \times \pi'_l$ değeri

	A	B	C
A	0,042	0,097	0,066
B	0,097	0,223	0,152
C	0,066	0,152	0,104

Uyum katsayılarının hesaplanmasına bu aşamadan sonra devam edilebilmesi için ağırlıklar hesaplanmalıdır. İlk olarak Eş. 2.2'ye göre hesaplanan kuadratik ağırlıklar Çizelge 3.7'de verilmiştir.

Çizelge 3.7. Kuadratik ağırlıklar

	A	B	C
A	1	0,75	0
B	0,75	1	0,75
C	0	0,75	1

Krippendorff'un α katsayısı hariç diğeri 4 katsayının formülünde yer alan ve gözlenen uyumu gösteren p_a , Çizelge 3.7'de verilen kuadratik ağırlıklarla p'_{kl} değeri çarpılarak hesaplanır. Bu yolla elde edilen p_a değeri Çizelge 3.8'deki tüm değeri toplamıdır.

Çizelge 3.8. $w_{kl} \times p'_{kl}$ değeri

	A	B	C
A	0,125	0,094	0,000
B	0,000	0,375	0,094
C	0,000	0,000	0,250

Buradan $p_a = 0,125 + 0,094 + 0 + 0 + 0,375 + 0,094 + 0 + 0 + 0,250 = 0,938$ olarak hesaplanır.

p_e değeriinin hesaplanması

Bu aşamada 2 değeriendirici için tüm uyum katsayılarının beklenen uyumu p_e değeri hesaplanır.

EK-3. (devam)İki değerdendiricili durum için örnek uygulama örneđi

Cohen'in κ katsayısını hesaplayabilmek için Eş. 2.2'ye göre $w_{kl} \times p'_{k+} \times p'_{+l}$ değerdlerinin hesaplanması gerekmektedir. Çizelge 3.9 $w_{kl} \times p'_{k+} \times p'_{+l}$ değerdlerini göstermektedir:

Çizelge 3.9. $w_{kl} \times p'_{k+} \times p'_{+l}$ değerdleri

	A	B	C
A	0,033	0,100	0,000
B	0,042	0,222	0,167
C	0,000	0,067	0,089

Cohen κ katsayısı için p_e değerdleri Çizelge 3.9'daki değerdlerin toplamıdır.

Scott'ın π katsayısını hesaplayabilmek için Eş. 2.37'ye göre $w_{kl} \times \pi'_k \times \pi'_l$ değerdlerinin hesaplanması gerekmektedir. Çizelge 3.10 $w_{kl} \times \pi'_k \times \pi'_l$ değerdlerini göstermektedir.

Çizelge 3.10. $w_{kl} \times \pi'_k \times \pi'_l$ d değerdleri

	A	B	C
A	0,042	0,073	0,000
B	0,073	0,223	0,114
C	0,000	0,114	0,104

Scott'ın π katsayısı için p_e değerdleri Çizelge 3.10'daki değerdlerin toplamıdır.

Brennan-Prediger B-P katsayısı için Eş. 2.50'ye göre p_e değerdleri hesaplanır.

$$p_e = \frac{1}{q^2} \sum_{k,l}^q w_{kl} = \frac{1}{9} (1 + 0,75 + \dots + 0,75 + 1) = 0,667$$

Gwet AC₂ için Eş. 2.61'e göre p_e değerdleri hesaplanır.

$$\begin{aligned}
 p_e &= \frac{T_w}{q(q-1)} \sum_{k=1}^q \pi'_k (1 - \pi'_k) \\
 &= \frac{6}{3 * 2} \cdot [0,206(1 - 0,206) + 0,472(1 - 0,472) + 0,322(1 - 0,322)] \\
 &= 0,631
 \end{aligned}$$

EK-3. (devam)İki değerlendiricili durum için örnek uygulama örneği

p_e değerleri Cohen'in κ 'sı için Eş. 2.5'e göre, Scott'ın π 'si için Eş. 2.37'ye göre, Brennan-Prediger BP katsayısı için Eş. 2.50'ye göre ve Gwet AC₂ için Eş. 2.61'e göre hesaplanmıştır ve aşağıdaki gibidir:

Çizelge 3.11. İki değerlendirici için uyum katsayıları için hesaplanan p_e değerleri

	κ	π	BP	AC ₂
p_e	0,719	0,743	0,667	0,631

Krippendorff'un α katsayısı için p_d ve p_e değerlerinin hesaplanması

Çizelge 3.1 verisi eksik skorlamalar içerdiğinden, 1., 6. ve 11. birim çalışmadan çıkartılarak hesaplanacaktır. Çizelge 3.12 Krippendorff'un α katsayısının hesaplanabilmesi için oluşturulan kontenjans tablosudur.

Çizelge 3.12. Eksik skorlamaları çıkartılmış Çizelge 3.1 verisi(n_{kl} değerleri)

Değerlendirici 1	Değerlendirici 2			Kayıp Skorlar	Toplam
	A	B	C		
A	1	1	0	0	2
B	0	3	1	0	4
C	0	0	2	0	2
Kayıp Skorlar	0	0	0	0	0
Toplam	1	4	3	0	8

Bu adımdan sonra yukarıdaki adımlar benzer şekilde takip edilir. Krippendorff'un α katsayısı için nispi oran p_{kl} değerleri Çizelge 3.13'te verilmiştir.

Çizelge 3.13. p_{kl} değerleri

Değerlendirici 1	Değerlendirici 2			Kayıp Skorlar	p_{k+}
	A	B	C		
A	0,125	0,125	0,000	0,000	0,250
B	0,000	0,375	0,125	0,000	0,500
C	0,000	0,000	0,250	0,000	0,250
Kayıp Skorlar	0,000	0,000	0,000	0,000	0,000
p_{+l}	0,125	0,500	0,375	0,000	1,000

EK-3. (devam)İki değerdendiricili durum için örnek uygulama örneđi

p_{kl} değerdleri hesaplandıktan sonra nispi oranlar hesaplanır. Nispi oranlar Çizelge 3.14'te verilmiştir.

Çizelge 3.14. p'_{kl} değerdleri

Değerdendirici 1	Değerdendirici 2			p'_{k+}	π_k
	A	B	C		
A	0,125	0,125	0,000	0,250	0,188
B	0,000	0,375	0,125	0,500	0,500
C	0,000	0,000	0,250	0,250	0,313
p'_{+l}	0,125	0,500	0,375		
π_l	0,188	0,500	0,313		

Bu adımdan sonra $\pi'_k \times \pi'_l$ değerdlerinin hesaplanması aşamasına geçilir. Çizelge 3.15, Krippendorff'un α değerdinin hesaplanması için $\pi'_k \times \pi'_l$ değerdlerini içermektedir.

Çizelge 3.15. $\pi'_k \times \pi'_l$ değerdleri

	A	B	C
A	0,035	0,094	0,059
B	0,094	0,250	0,156
C	0,059	0,156	0,098

Bu adımdan sonra kuadratik ağırlıklar hesaplanmalıdır. Eş. 2.1'e göre hesaplanan Kuadratik ağırlıklar Çizelge 3.7'deki gibidir. Kuadratik ağırlıklarla p'_{kl} değerdleri çarpılarak gözlenen uyum değeri olan p_a değeri hesaplanır. p_a değeri Çizelge 3.8'dek tüm değerdlerin toplamıdır ve Krippendorff'un α katsayısı için de aynı değere sahiptir.

Krippendorff'un α katsayısını hesaplayabilmek için $w_{kl} \times \pi'_k \times \pi'_l$ değerdlerini hesaplamak gerekir. Çizelge 3.16 bu değerdleri göstermektedir:

Çizelge 3.16. $w_{kl} \times \pi'_k \times \pi'_l$ d değerdleri

	A	B	C
A	0,125	0,094	0,000
B	0,000	0,375	0,094
C	0,000	0,000	0,250

EK-3. (devam)İki değerdendiricili durum için örnek uygulama örneđi

Bu adımdan sonra Krippendorff'un α katsayısı için beklenen uyumu p_e değeri hesaplanır. p_e Eş.2.55'e göre hesaplanmıştır. Krippendorff'un α katsayısı için $p_e = 0,758$ olarak ve Krippendorff'un α katsayısı için p_a değeri Eş. 2.55'ten $p_e = 0,941$ olarak hesaplanmıştır.

Çizelge 3.16, κ , π , BP, α ve AC2 katsayılarıyla bunların ağırlıklandırılmış ve ağırlıklandırılmamış versiyonlarındaki yüzde uyumu göstermektedir. Bu katsayıların ağırlıklandırılmış versiyonları *kuadratik ağırlıkları* kullanmaktadır.

- Cohen'in Kappa'sı, $\hat{\kappa}_C = (0,938 - 0,719)/(1 - 0,719) = 0,777$
- Scott'un π 'si, $\hat{\kappa}_S = (0,938 - 0,743)/(1 - 0,743) = 0,7569$
- Brennan-Prediger Katsayısı, $\hat{\kappa}_G = (0,938 - 0,667)/(1 - 0,667) = 0,813$
- Krippendorff'un α 'sı, $\hat{\kappa}_S = (0,941 - 0,758)/(1 - 0,758) = 0,758$
- Gwet'in AC₂'si, $\hat{\kappa}_G = (0,938 - 0,631)/(1 - 0,631) = 0,831$

Çizelge 3.17. Çizelge 3.1'deki veriden elde edilen kuadratik ağırlıklarla ağırlıklandırılmış ve ağırlıklandırılmamış katsayılar

Uyum Katsayıları	Ağırlıklandırılmamış	Ağırlıklandırılmış	Standart Hata
Cohen'in κ 'sı	0,617	0,777	0,137
Scott'un π 'si	0,604	0,758	0,108
Brennan-Prediger BP	0,625	0,813	0,160
Krippendorff'un α 'sı	0,620	0,758	0,163
Gwet'in AC ₂ 'si	0,635	0,831	0,115

Aynı adımlar basit sıralı ağırlıklar kullanıldığında da benzer şekilde olup, basit sıralı ağırlıklar kullanılarak hesaplanmış 2 değerdendirici için uyum katsayıları aşağıdaki tabloda verilmiştir:

Çizelge 3.18. Üç kategori için basit sıralı ağırlıklar

	A	B	C
A	1,000	0,667	0,000
B	0,667	1,000	0,667
C	0,000	0,667	1,000

EK-3. (devam)İki değerdendiricili durum için örnek uygulama örneđi

Çizelge 3.19: Çizelge 3.1’deki veriden elde edilen basit sıralı ağırlıklarla ağırlıklandırılmış ve ağırlıklandırılmamış katsayılar

Uyum Katsayıları	Ağırlıklandırılmamış	Ağırlıklandırılmış	Standart Hata
Cohen’in κ ’sı	0,619	0,741	0,157
Scott’un π ’si	0,604	0,721	0,130
Brennan-Prediger BP	0,625	0,775	0,180
Krippendorff’un α ’sı	0,620	0,725	0,181
Gwet’in AC_2 ’si	0,635	0,794	0,138

EK-4. Üç veya daha fazla değerlendiricili durum için uyum hesaplaması örnek uygulama

Bu örnekte, 5 değerlendiricili uyum katsayısının hesaplanmasını göstermek için, Çizelge 4.1'in skorlanmış verisini dikkate alındı[Gwet, 2014;s.90]. “0,5”, “1,0”, “1,5”, “2,0” ve “2,5” gibi 5 skordan biri L, K, W ve B olarak etiketlenen 4 değerlendirici tarafından deneklere atandı. Skorlamalar sıralı tipte olduğundan, analiz için ağırlıklandırılmış bir katsayı kullanmak gerekmektedir. Bu veri, aynı zamanda eksik(kayıp) skorlamalar içermektedir. Örneğin, “a.logan” isimli denek sadece K ve W değerlendiricileri tarafından skorlanmıştır.

Çizelge 4.1. 16 deneğin 4 değerlendirici tarafından skorlamaları

Subject	L	K	W	B
a.lycia	1	1,5	1	
a.milbe	2	2	2	2
a.hegon	0,5	1	1,5	1,5
a.oslar	1	1	1	1
a.viali	1	1	1	1,5
a.logan		1	2,5	
a.numit	2,5	2,5	2,5	2,5
a.saraa	1	1		1
a.sarat		1	2	1
a.mormo	1	1	0,5	1
a.celti	1,5	1,5	1,5	1,5
a.clyto	1	1,5	1	
a.hiann	1	1	1,5	
b.phile	1	2	2,5	2
b.alask		1	1,5	1
b.taad	0,5	0,5	0,5	0,5

Çizelge 4.1'in skollama verisini ele alınsın. Bu veri seti ağırlıklandırılmış ve ağırlıklandırılmamış uyum katsayılarının hesaplanmasında kullanılacaktır. Krippendorff'un α katsayısının yüzde şans uyumu p_a 'nın formülü diğer katsayılara göre farklı hesaplandığından, ilk olarak Conger'in κ 'sı, B-P katsayısı, Fleiss'in κ 'sı ve Gwet AC₂ uyum katsayılarının yüzde şans uyumu p_a değerleri hesaplanacaktır.

Eş. 2.68'e göre yüzde şans uyumu p_a , r_{ik} ve r^*_{ik} değerlerinin bulunmasıyla hesaplanır. Çizelge 4.2 ve Çizelge 4.3'te bu değerler verilmiştir.

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Yüzde uyum(gözlenen uyum) p_a 'nin hesaplanması

Çizelge 4.2. r_{ik} değerdleri

Birimler	0,5	1	1,5	2	2,5	r_i
A_lycia	0	2	1	0	0	3
A_milbe	0	0	0	4	0	4
A_hegon	1	1	2	0	0	4
A_oslar	0	4	0	0	0	4
A_viali	0	4	0	0	0	4
A_logan	0	1	0	0	1	2
A_numit	0	0	0	0	4	4
A_saraa	0	3	0	0	0	3
A_sarat	0	2	0	1	0	3
A_mormo	1	3	0	0	0	4
A_celti	0	0	4	0	0	4
A_clyto	0	2	1	0	0	3
A_hiann	0	2	1	0	0	3
B_phile	0	1	0	2	1	4
B_alask	0	2	1	0	0	3
B_taad	4	0	0	0	0	4
					$\bar{r}$	3,5

Çizelge 4.3: $r *_{ik}$ değerdleri

Birimler	0,5	1	1,5	2	2,5
A_lycia	2,625	2,9375	2,875	2,4375	1,625
A_milbe	1,75	3	3,75	4	3,75
A_hegon	3,4375	3,8125	3,6875	3,0625	1,9375
A_oslar	3,75	4	3,75	3	1,75
A_viali	3,75	4	3,75	3	1,75
A_logan	0,9375	1,4375	1,6875	1,6875	1,4375
A_numit	0	1,75	3	3,75	4
A_saraa	2,8125	3	2,8125	2,25	1,3125
A_sarat	2,3125	2,75	2,8125	2,5	1,8125
A_mormo	3,8125	3,9375	3,5625	2,6875	1,3125
A_celti	3	3,75	4	3,75	3
A_clyto	2,625	2,9375	2,875	2,4375	1,625
A_hiann	2,625	2,9375	2,875	2,4375	1,625
B_phile	1,8125	2,9375	3,5625	3,6875	3,3125
B_alask	2,625	2,9375	2,875	2,4375	1,625
B_taad	4	3,75	3	1,75	0

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Bu aşamadan sonra Çizelge 4.4'te $\frac{r_{ik}(r_{ik}^*-1)}{r_i(r_{i-1})}$ değerdlerinin hesaplaması yapılmıştır.

Çizelge 4.4. $\frac{r_{ik}(r_{ik}^*-1)}{r_i(r_{i-1})}$ değerdleri

Birimler	0,5	1	1,5	2	2,5
A_lycia	0,000	0,646	0,313	0,000	0,000
A_milbe	0,000	0,000	0,000	1,000	0,000
A_hegon	0,203	0,234	0,448	0,000	0,000
A_oslar	0,000	1,000	0,000	0,000	0,000
A_viali	0,000	0,734	0,234	0,000	0,000
A_logan	0,000	0,219	0,000	0,000	0,219
A_numit	0,000	0,000	0,000	0,000	1,000
A_saraa	0,000	1,000	0,000	0,000	0,000
A_sarat	0,000	0,583	0,000	0,250	0,000
A_mormo	0,234	0,734	0,000	0,000	0,000
A_celti	0,000	0,000	1,000	0,000	0,000
A_clyto	0,000	0,646	0,313	0,000	0,000
A_hiann	0,000	0,646	0,313	0,000	0,000
B_phile	0,000	0,161	0,000	0,448	0,193
B_alask	0,000	0,646	0,313	0,000	0,000
B_taad	1,000	0,000	0,000	0,000	0,000

Yukarı tablolar elde edildikten sonra ilgili 4 uyum katsayısı için yüzde şans uyumu p_a değeri aşağıdaki gibi bulunmuştur.

$$p_a = 1/16 \times (0 + 0,646 + 0,313 + \dots + 0 + 0 + 0) = 0,921$$

Krippendorff'un α katsayısının p_a değerdinin hesaplanması

Çizelge 4.5'te, Eş. 2.99'da verilen p'_a değerdinin hesaplanmıştır.

Çizelge 4.5. $\frac{r_{ik}(r_{ik}^*-1)}{r(r_i-1)}$ değerdleri

Birimler	0,5	1	1,5	2	2,5
A_lycia	0,000	0,554	0,268	0,000	0,000
A_milbe	0,000	0,000	0,000	1,143	0,000
A_hegon	0,232	0,268	0,512	0,000	0,000
A_oslar	0,000	1,143	0,000	0,000	0,000

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Çizelge 4.5.(devam) $\frac{r_{ik}(r_{ik}^*-1)}{r(r_i-1)}$ değerdleri

A_viali	0,000	0,839	0,268	0,000	0,000
A_logan	0,000	0,125	0,000	0,000	0,125
A_numit	0,000	0,000	0,000	0,000	1,143
A_saraa	0,000	0,857	0,000	0,000	0,000
A_sarat	0,000	0,500	0,000	0,214	0,000
A_mormo	0,268	0,839	0,000	0,000	0,000
A_celti	0,000	0,000	1,143	0,000	0,000
A_clyto	0,000	0,554	0,268	0,000	0,000
A_hiann	0,000	0,554	0,268	0,000	0,000
B_phile	0,000	0,185	0,000	0,512	0,220
B_alask	0,000	0,554	0,268	0,000	0,000
B_taad	1,143	0,000	0,000	0,000	0,000

Buradan Krippendorff'un α katsayısı için p_a yüzde şans uyumu aşağıdaki gibi hesaplanır:

$$p'_a = \frac{1}{n'} \sum_{i=1}^{n'} \sum_{k=1}^q \frac{r_{ik}(r_{ik}^*-1)}{r(r_i-1)} = \frac{1}{16} * (0 + 0,554 + 0,268 + \dots + 0 + 0 + 0) = 0,935$$

Çizelge 4.2'ye göre $\bar{r} = 3,5$ olarak hesaplandığından Krippendorff'un α katsayısı p_a değeri

Eş. 2.99'a göre aşağıdaki gibi hesaplanır:

$$p_a = (1 - \varepsilon_n)p'_a + \varepsilon_n = \left[1 - \left(\frac{1}{16} * 3,5\right)\right] \times 0,935 + \left[1 - \left(\frac{1}{16} * 3,5\right)\right] = 0,936$$

Şansa bağılı yüzde uyum(beklenen uyum) p_e 'nin hesaplanması

Conger'in κ 'sı

Çizelge 4.6 ve 4.7'de n_{gk} ve p_{gk} değerdleri, s_{kl}^2 değerdlerinin hesaplanabilmesi için bulunması gereken değerdlerdir.

Çizelge 4.6. n_{gk} değerdleri

Değerdlendiriciler	0,5	1	1,5	2	2,5	n_g
L	2	8	1	1	1	13
K	1	9	3	2	1	16

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Çizelge 4.6. (devam) n_{gk} değerdleri

W	2	4	4	2	3	15
B	1	5	3	2	1	12
Ortalama	1,5	6,5	2,8	1,8	1,5	14

n_{gk} değerdleri hesaplandıktan sonra nispi oranlar hesaplanır. Nispi oranlar Çizelge 4.7’de verilmiştir.

Çizelge 4.7. p_{gk} değerdleri

Değerdendiriciler	0,5	1	1,5	2	2,5
L	0,154	0,615	0,077	0,077	0,077
K	0,063	0,563	0,188	0,125	0,063
W	0,133	0,267	0,267	0,133	0,200
B	0,083	0,417	0,250	0,167	0,083
Ortalama	0,108	0,465	0,195	0,125	0,106

Bu adımdan sonra s_{kl}^2 değerdleri Çizelge 4.8’de hesaplanmıştır.

Çizelge 4.8. s_{kl}^2 değerdleri

	0,5	1	1,5	2	2,5
0,5	0,002	0,000	-0,002	-0,001	0,001
1	0,000	0,025	-0,012	-0,004	-0,009
1,5	-0,002	-0,012	0,007	0,003	0,003
2	-0,001	-0,004	0,003	0,001	0,000
2,5	0,001	-0,009	0,003	0,000	0,004

Bir sonraki adımda $\bar{p}_{+k}\bar{p}_{+l}$ çarpımının hesaplanması gerekmektedir.

Çizelge 4.9. $\bar{p}_{+k}\bar{p}_{+l}$ değerdleri

	0,5	1	1,5	2	2,5
0,5	0,011	0,047	0,016	0,006	0,000
1	0,047	0,210	0,088	0,044	0,022
1,5	0,016	0,088	0,036	0,022	0,015
2	0,006	0,044	0,022	0,015	0,012
2,5	0,000	0,022	0,015	0,012	0,010

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Bir sonraki adımda $w_{kl}(\bar{p}_{+k}\bar{p}_{+l} - s_{kl}^2/r)$ çarpımının hesaplanması gerekmektedir. İlk olarak w_{kl} değeri için kuadratik ağırlıklar kullanılmıştır. Çizelge 4.10 kuadratik ağırlıkları göstermektedir:

Çizelge 4.10. Kuadratik ağırlıklar(w_{kl})

	0,5	1	1,5	2	2,5
0,5	1,000	0,938	0,750	0,438	0,000
1	0,938	1,000	0,938	0,750	0,438
1,5	0,750	0,938	1,000	0,938	0,750
2	0,438	0,750	0,938	1,000	0,938
2,5	0,000	0,438	0,750	0,938	1,000

Çizelge 4.11. $w_{kl}(\bar{p}_{+k}\bar{p}_{+l} - s_{kl}^2/r)$ değeri

	0,5	1	1,5	2	2,5
0,5	0,011	0,047	0,016	0,006	0,000
1	0,047	0,210	0,088	0,044	0,022
1,5	0,016	0,088	0,036	0,022	0,015
2	0,006	0,044	0,022	0,015	0,012
2,5	0,000	0,022	0,015	0,012	0,010

Buradan Conger'in κ 'sı için şansa bağlı yüzde uyum(beklenen uyum) p_e aşağıdaki gibi hesaplanır:

$$p_e = \sum_{k,l} w_{kl}(\bar{p}_{+k}\bar{p}_{+l} - s_{kl}^2/r)$$

$$= (0,011 + 0,047 + 0,016 + \dots + 0,015 + 0,015 + 0,010) = 0,831$$

Brennan-Prediger'in B-P katsayısı

Eş. 2.82'ye göre Brennan-Prediger'in B-P katsayısının p_e değeri Çizelge 4.10'daki ağırlıklar kullanılarak aşağıdaki gibi hesaplanır:

$$p_e = \frac{1}{q^2} \sum_{k,l} w_{kl} = \frac{1}{25} \times (1 + 0,938 + 0,750 + \dots + 0,750 + 0,938 + 1) = 0,750$$

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Fleiss'in κ 'sı

Eş. 2.88'e göre Fleiss'in κ katsayısının p_e değeri Çizelge 4.10'daki ağırlıklar kullanılarak aşağıdaki gibi hesaplanır:

Çizelge 4.12. Fleiss'in κ 'sı için $w_{kl}\pi_k\pi_l$ değeri

	0,5	1	1,5	2	2,5
0,5	0,009	0,043	0,014	0,005	0,000
1	0,043	0,240	0,088	0,042	0,023
1,5	0,014	0,088	0,037	0,021	0,016
2	0,005	0,042	0,021	0,013	0,012
2,5	0,000	0,023	0,016	0,012	0,012

$$p_e = \sum_{k,l} w_{kl} \pi_k \pi_l = (0,009 + 0,043 + 0,014 + \dots + 0,016 + 0,012 + 0,012) = 0,838$$

Krippendorff'un α 'sı

Eş. 2.99'a göre Krippendorff'un α katsayısının p_e değeri aşağıdaki gibi hesaplanmıştır.

Çizelge 4.13. $\pi_k\pi_l$ değeri

Birimler	0,5	1	1,5	2	2,5
A_lycia	0,000	0,571	0,286	0,000	0,000
A_milbe	0,000	0,000	0,000	1,143	0,000
A_hegon	0,286	0,286	0,571	0,000	0,000
A_oslar	0,000	1,143	0,000	0,000	0,000
A_viali	0,000	0,857	0,286	0,000	0,000
A_logan	0,000	0,286	0,000	0,000	0,286
A_numit	0,000	0,000	0,000	0,000	1,143
A_saraa	0,000	0,857	0,000	0,000	0,000
A_sarat	0,000	0,571	0,000	0,286	0,000
A_mormo	0,286	0,857	0,000	0,000	0,000
A_celti	0,000	0,000	1,143	0,000	0,000
A_clyto	0,000	0,571	0,286	0,000	0,000
A_hiann	0,000	0,571	0,286	0,000	0,000
B_phile	0,000	0,286	0,000	0,571	0,286
B_alask	0,000	0,571	0,286	0,000	0,000
B_taad	1,143	0,000	0,000	0,000	0,000
π_k	0,107	0,464	0,196	0,125	0,107

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Çizelge 4.14 Krippendorff'un α sı için $w_{kl}\pi_k\pi_l$ değerdleri

	0,5	1	1,5	2	2,5
0,5	0,011	0,047	0,016	0,006	0,000
1	0,047	0,216	0,085	0,044	0,022
1,5	0,016	0,085	0,039	0,023	0,016
2	0,006	0,044	0,023	0,016	0,013
2,5	0,000	0,022	0,016	0,013	0,011

Buradan Krippendorff'un α katsayısına ilişkin p_e şansa bağılı yüzde uyum aşğıdaki gibi hesaplanmıştır:

$$p_e = \sum_{k,l} w_{kl}\pi_k\pi_l = (0,011 + 0,047 + 0,016 + \dots + 0,016 + 0,013 + 0,011 = 0,834$$

Gwet AC₂'si

Eş. 2.111'e göre Gwet'in AC₂ katsayısının p_e değeri aşğıdaki gibi hesaplanmıştır. $\pi_k(1 - \pi_k)$ değerdleri Çizelge 4.14'ten elde edilmiştir.

Çizelge 4.15. $\pi_k(1 - \pi_k)$ değerdleri

Birimler	0,5	1	1,5	2	2,5
A_lycia	0,000	0,667	0,333	0,000	0,000
A_milbe	0,000	0,000	0,000	1,000	0,000
A_hegon	0,250	0,250	0,500	0,000	0,000
A_oslar	0,000	1,000	0,000	0,000	0,000
A_viali	0,000	0,750	0,250	0,000	0,000
A_logan	0,000	0,500	0,000	0,000	0,500
A_numit	0,000	0,000	0,000	0,000	1,000
A_saraa	0,000	1,000	0,000	0,000	0,000
A_sarat	0,000	0,667	0,000	0,333	0,000
A_mormo	0,250	0,750	0,000	0,000	0,000
A_celti	0,000	0,000	1,000	0,000	0,000
A_clyto	0,000	0,667	0,333	0,000	0,000
A_hiann	0,000	0,667	0,333	0,000	0,000
B_phile	0,000	0,250	0,000	0,500	0,250
B_alask	0,000	0,667	0,333	0,000	0,000
B_taad	1,000	0,000	0,000	0,000	0,000
π_k	0,094	0,490	0,193	0,115	0,109

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Buradan,

$$p_e = \frac{T_w}{q(q-1)} \sum_{k,l} \pi_k(1 - \pi_k) = \frac{18,750}{5 \cdot 4} \cdot (0,094(1 - 0,094) + \dots + 0,109(1 - 0,109))$$

$$= 0,646$$

olarak bulunur.

Krippendorff'un ağırlıklandırılmış yüzde uyumu ise $p_a = 0,935$ olarak verilmiştir. Ancak, tüm uyum katsayıları farklı p_e (yüzde şansa bağılı uyum) değerlerine sahiptir. Bu uyum katsayıları aşağıdaki gibidir:

- Conger'in π 'sı, $\hat{\kappa}_C = (0,921 - 0,831)/(1 - 0,831) = 0,529$
- Brennan-Prediger B-P katsayısı, $\hat{\kappa}_G = (0,921 - 0,75)/(1 - 0,75) = 0,682$
- Fleiss'in κ 'sı, $\hat{\kappa}_C = (0,921 - 0,838)/(1 - 0,838) = 0,511$
- Krippendorff'un α 'sı, $\hat{\alpha}_K = (0,936 - 0,834)/(1 - 0,834) = 0,618$
- Gwet'in AC₂'si, $\hat{\kappa}_G = (0,921 - 0,646)/(1 - 0,646) = 0,775$

Çizelge 4.16. Çizelge 4.1 verisi için hesaplanan uyum katsayılarının ağırlıklandırılmış ve ağırlıklandırılmamış değerleri

	Ağırlıklandırılmamış	Ağırlıklandırılmış	StdHata
Conger's Kappa	0,38941	0,52897	0,16452
Gwet's AC1	0,47141	0,77550	0,10902
Fleiss' Kappa	0,36529	0,51065	0,22575
Krippendorff's Alpha	0,43793	0,61796	0,16359
Brenann-Prediger	0,45313	0,68229	0,14246

Çizelge 4.16. Çizelge 4.1 verisi için hesaplanan uyum katsayılarının ağırlıklandırılmış hesaplamaları için standart hataları verilmiştir.

Basit sıralı ağırlıklarla da hesaplama yapılmıştır. Çizelge 4.17'de basit sıralı ağırlıklar verilmiştir.

EK-4. (devam)Üç değerdendiricili durum için uyum hesaplaması örnek uygulama

Çizelge 4.17. Basit sıralı ağırlıklar(w_{kl})

	0,5	1	1,5	2	2,5
0,5	1	0,9	0,7	0,4	0
1	0,9	1	0,9	0,7	0,4
1,5	0,7	0,9	1	0,9	0,7
2	0,4	0,7	0,9	1	0,9
2,5	0	0,4	0,7	0,9	1

Çizelge 4.18. Çizelge 4.1 verisi için hesaplanan uyum katsayılarının ağırlıklandırılmış ve ağırlıklandırılmamış değerdleri

	Ağırlıklandırılmamış	Ağırlıklandırılmış	StdHata
Conger's Kappa	0,38941	0,50869	0,15100
Gwet's AC1	0,47141	0,74482	0,10827
Fleiss' Kappa	0,36529	0,48952	0,21197
Krippendorff's Alpha	0,43793	0,59139	0,15892
Brenann-Prediger	0,45313	0,65402	0,13630

Her iki ağırlık kümesi için de yapılan hesaplamalarda, Gwet AC₂ katsayısı sırasıyla Conger'in, Fleiss'in ve Krippendorff'dan yüksek olan Brennan-Prediger katsayısından da büyüktür. Bu Conger, Fleiss ve Krippendorff'un p_e : şansa bağılı uyum olduğundan fazla hesaplama eğilimlerinden kaynaklanmaktadır.

EK-5. Landis Ve Koch (1977b) verisi

#	A	B	C	D	E	F	G	#	A	B	C	D	E	F	G	#	A	B	C	D	E	F	G
1	4	3	4	2	3	3	3	48	3	3	3	2	3	2	3	94	1	1	2	1	2	1	1
2	1	1	1	1	1	1	1	49	3	2	2	2	2	1	1	95	3	3	3	2	4	3	3
3	3	3	3	3	3	3	3	51	2	3	2	2	2	2	2	96	4	3	1	1	2	1	2
4	4	3	3	4	3	3	3	52	3	3	3	4	3	2	3	98	4	3	3	4	4	3	3
5	3	3	3	3	3	3	3	53	4	3	3	3	3	5	3	99	1	2	2	1	2	1	2
6	2	1	2	1	1	1	1	54	3	3	2	2	4	2	3	100	3	3	3	2	4	2	3
7	1	1	1	1	2	1	1	55	3	3	3	3	3	2	3	101	4	4	3	4	4	3	4
8	3	3	2	3	2	2	3	56	2	2	2	1	2	2	2	102	3	3	2	2	3	3	3
9	2	2	2	2	3	1	2	57	2	3	2	2	3	1	3	103	1	1	1	1	1	1	1
10	1	1	1	1	2	1	1	58	1	1	1	1	1	1	1	104	2	3	2	2	4	1	2
11	5	5	5	4	5	5	5	59	3	3	3	3	3	3	3	105	3	3	3	3	3	2	3
12	1	1	1	1	2	1	1	60	1	1	2	1	1	1	1	106	2	3	1	1	3	1	1
13	3	3	3	2	3	3	3	61	1	3	2	1	2	1	1	107	3	3	2	2	3	2	3
15	2	2	2	1	1	1	2	62	4	3	3	3	3	2	3	108	3	3	2	2	3	1	3
16	4	3	3	2	3	2	3	63	1	3	2	2	2	1	2	110	2	2	1	1	2	1	1
17	3	3	2	3	3	3	3	64	2	3	2	2	3	2	3	111	1	1	1	1	2	1	1
18	2	3	2	2	2	1	3	65	4	3	3	3	3	3	3	112	3	3	2	2	2	2	3
19	2	1	2	1	2	1	1	66	3	3	3	4	3	2	4	113	3	3	2	2	2	1	2
22	2	3	2	2	2	1	3	67	1	1	1	1	1	1	1	114	2	3	1	1	2	1	1
23	1	1	2	1	1	1	1	68	2	3	2	2	3	2	2	115	3	3	2	2	3	2	3
24	4	3	3	4	3	3	3	69	3	3	2	3	3	1	3	116	1	1	1	1	2	1	1
25	1	1	2	1	2	1	1	70	1	1	1	1	1	1	1	117	3	3	3	2	3	2	3
26	1	1	1	1	1	1	1	71	4	3	3	3	3	3	3	118	3	3	2	2	3	1	3
27	2	1	2	2	2	1	2	72	3	3	3	2	3	1	3	119	1	1	1	1	2	1	1
28	4	4	4	2	4	3	3	73	3	3	3	3	3	2	3	120	1	1	1	1	1	1	1
29	3	3	3	2	3	2	3	74	4	3	1	3	3	2	3	121	2	2	1	1	2	1	2
30	3	3	3	3	3	2	3	76	1	2	1	1	1	1	1	122	5	3	4	2	3	4	3
31	1	1	1	1	1	1	1	77	2	2	1	2	2	1	2	123	4	3	4	2	4	1	3
32	4	3	3	3	3	2	3	78	2	3	2	1	3	2	2	124	1	1	1	1	2	1	1
33	3	3	3	3	3	3	3	79	2	1	1	2	1	1	1	126	2	3	1	1	2	1	2
34	1	1	1	1	1	1	1	80	4	4	3	2	4	1	3								
35	3	3	3	2	3	1	3	81	1	1	1	1	1	1	1								
36	2	2	2	2	3	1	2	82	4	4	3	3	4	3	3								
37	3	3	2	2	3	1	3	83	5	5	1	4	5	5	4								
38	5	3	3	3	4	1	3	84	2	3	2	2	2	1	2								
39	2	1	1	1	2	1	1	85	4	4	4	2	5	1	3								
40	3	3	2	2	3	1	3	86	3	3	2	3	3	3	3								
41	3	3	3	3	3	2	3	87	4	3	3	3	3	3	3								
42	5	5	5	5	5	5	5	88	4	2	3	2	3	2	3								
43	5	3	3	2	3	2	3	89	2	3	2	2	4	1	3								
44	3	2	2	2	2	1	2	90	3	3	3	2	4	2	3								
45	1	1	1	1	2	1	1	91	3	3	2	1	3	2	2								
46	2	3	1	2	3	1	3	92	4	4	3	2	4	1	3								
47	4	4	4	3	3	3	3	93	3	3	2	2	3	2	2								

EK-6. Stacklefish verisi

	Değerlendirici 1	Değerlendirici 2	Değerlendirici 3	Değerlendirici 4
1	5	5	5	5
2	3	1	3	1
3	5	5	5	5
4	3	1	3	1
5	5	5	4	5
6	1	2	3	3
7	3	1	1	1
8	1	1	1	3
9	3	3	4	4
10	1	3	1	1
11	5	5	5	5
12	1	1	1	1
13	1	1	1	1
14	1	1	1	1
15	3	3	4	3
16	1	3	3	4
17	4	4	5	5
18	5	5	5	5
19	5	3	3	3
20	3	2	3	3
21	5	3	5	5
22	3	3	3	4
23	1	1	1	1
24	1	1	1	1
25	1	1	3	3
26	3	3	3	1
27	3	3	1	1
28	3	3	1	1
29	3	3	2	5

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : TUTALAR, Nilden
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 18.02.1984, Ankara
 Medeni hali : Bekâr
 Telefon : 0 (534) 747 56 84
 Faks : -
 e-mail : nildentutalar@windowslive.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Yüksek lisans	Gazi Üniversitesi /İstatistik	Devam Ediyor
Lisans	Gazi Üniversitesi /İstatistik	2006
Lise	Çağrıbey Anadolu Lisesi/Fen Bilimleri	2002

İş Deneyimi

Yıl	Yer	Görev
2011-2013	Pidosan A.Ş.	Kalite Müdürü

Yabancı Dil

İngilizce/YDS 82.5/TOEFL IBT 70

Yayınlar

“Comparison of Inter-Rater Reliability Measurement Methods For Ordinal Scaled Scores”,
 The 8th conference of the Eastern Mediterranean Region of the International Biometric
 Society (EMR-IBS), May 11-15,2015

Hobiler

Müzik, doğa yürüyüşleri



GAZİ GELECEKTİR...