

T.C.

UKUROVA NİVERSİTESİ

SAĐLIK BİLİMLERİ ENSTİTÜSÜ

BİYOİSTATİSTİK ANABİLİM DALI

**GÜVENİRLİK ALIŞMALARINDA ÖRNEK BÜYÜKLÜĐÜ
HESAPLAMA YAKLAŞIMLARI**

Sevin Püren YÜCEL

BİYOİSTATİSTİK TEZLİ YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

DANIŞMANI

Prof. Dr. Z. Nazan ALPARSLAN

ADANA-2017

T.C.

ÇUKUROVA ÜNİVERSİTESİ

SAĞLIK BİLİMLERİ ENSTİTÜSÜ

BİYOİSTATİSTİK ANABİLİM DALI

GÜVENİRLİK ÇALIŞMALARINDA ÖRNEK BÜYÜKLÜĞÜ HESAPLAMA YAKLAŞIMLARI

Sevinç Püren YÜCEL

BİYOİSTATİSTİK TEZLİ YÜKSEK LİSANS PROGRAMI

YÜKSEK LİSANS TEZİ

DANIŞMANI

Prof. Dr. Z. Nazan ALPARSLAN

Bu tez Çukurova Üniversitesi Bilimsel Araştırma Proje Birimi tarafından BAP TYL-
2017-7916 nolu proje olarak desteklenmiştir.

ADANA-2017

KABUL VE ONAY

Biyoistatistik Anabilim Dalı

Yüksek Lisans Programı Çerçevesinde yürütölmüş olan

“Güvenirlik Çalışmalarında Örnek Büyüklüğü Hesaplanma Yaklaşımları”
adlı çalışma, aşağıdaki jüri tarafından **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Tarihi: / /

TEZ SINAV JÜRİSİ

Başkan

Üye

Üye

Yukarıdaki Tez, Yönetim Kurulunun / / tarih ve sayılı kararı ile
kabul edilmiştir.

Prof.Dr. Behice DURGUN

Sağlık Bilimleri Enstitü Müdürü

T.C. ÇUKUROVA ÜNİVERSİTESİ SAĞLIK BİLİMLERİ
ENSTİTÜSÜ YÜKSEK LİSANS / DOKTORA TEZ ÇALIŞMASI
ETİK BEYANI

Çukurova Üniversitesi Bilimsel Araştırma ve Yayın Etiği Yönergesini okuduğumu ve anladığımı ve Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik kurallarına uygun olarak sunduğumu,
 - Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde ve ortaya çıkan sonuçlarda herhangi bir değişiklik yapmadığımı,
 - Tez olarak sunduğum bu çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda bu konuda hakkımda yapılacak tüm yasal işlemleri ve aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. ../../2017

Sevinç Püren YÜCEL

Kayıtlı olunan Program : Biyoistatistik Tezli YL
Tezin Konusu : Güvenirlik Çalışmalarında Örnek Büyüklüğü Hesaplama Yaklaşımları

Tezin Türü : Yüksek Lisans : ☒ Doktora: ☐

Danışmanın Adı-Soyadı : Prof. Dr. Z.Nazan ALPARSLAN

Danışmanın İletişim Bilgileri

Telefon : 0 (322) 338 60 84 - 3094

E-Posta : nazan@cu.edu.tr

Öğrencinin İletişim Bilgileri

Telefon : 0 (507) 770 85 09

E-Posta : sevincpurenyucel@gmail.com

Adresi : Çukurova Üni. Tıp Fakültesi
Biyoistatistik AD. ADANA

**Bu belgenin Lisansüstü eğitim tezleri savunmaya alınmadan önce öğrenci tarafından doldurulup imzalanarak Enstitü Müdürlüğüne teslim edilmesi gerekmektedir.*

TEŞEKKÜR

Yüksek lisans eğitimim ve tezim süresince tecrübesi ve anlayışıyla bana yol gösteren, benden ilgisini ve desteğini esirgemeyen değerli tez danışmanım Prof. Dr. Z. Nazan ALPARSLAN'a, ilgilerinden dolayı Prof. Dr. Gülşah SEYDAOĞLU'na, Doç. Dr. Yaşar SERTDEMİR'e ve Yrd. Doç. Dr. İlker ÜNAL'a, değerli katkılarından dolayı Mersin Üniversitesi öğretim üyesi Doç. Dr. Gülhan OREKİCİ TEMEL'e ve Yrd. Doç. Dr. Semra ERDOĞAN'a, çalışmanın gerçekleşmesi için verileri sağlayan Dr. Nihan CÜZDAN COŞKUN'a ve Arş. Gör. İpek KÖSE'ye; güler yüzleri, içtenlikleri ve destekleri için değerli arkadaşlarım Arş. Gör. Betül DAĞOĞLU ve Arş. Gör. Nazlı TOTİK'e, öğrenim hayatım boyunca destekleri ve anlayışlarından dolayı AİLEM'e içtenlikle teşekkür ederim.

Sevinç Püren YÜCEL

İÇİNDEKİLER

KABUL VE ONAY	ii
ETİK BEYANI	iii
TEŞEKKÜR	iv
İÇİNDEKİLER	v
ÇİZELGELER DİZİNİ	vii
ŞEKİLLER DİZİNİ	ix
SİMGELER ve KISALTMALAR DİZİNİ	xi
ÖZET	xii
ABSTRACT	xiii
1.GİRİŞ	1
2.GENEL BİLGİLER	3
2.1.Ölçeklerde Geçerlik ve Güvenirlik	3
2.2.Klasik Test Teorisi	4
2.2.1.Klasik Test Teorisinin Tanımlanması	4
2.2.2.Klasik Test Teorisi Modelleri	6
2.2.3.Geçerlik ve Güvenirlik Çalışmalarında Hatalar	7
2.3.Geçerlik	8
2.3.1.Geçerlik Çeşitleri	9
2.4.Güvenirlik	15
2.5.Güvenirliğin Belirlenmesinde Kullanılan Yöntemler	16
2.5.1.Test-Tekrar Test Yöntemi	17
2.5.2.Paralel veya Alternatif Testler Yöntemi	19
2.5.3.Testin İki Yarıya Bölünmesi Yöntemi	20
2.6.Bir Güvenirlik Ölçütü olarak Cronbach'ın Alfa'sı	23
2.6.1.Alfa Katsayısının Yorumlanması	26
2.6.2.Alfa Katsayısı için Güven Aralığı	27
2.6.3.Güvenirlik Çalışmalarında Alfa Katsayısına Bağlı Örnek Büyüklüğü	29
2.6.3.1.İstenilen Doğruluk için Örnek Büyüklüğü Hesaplama Yaklaşımı	30
2.6.3.2.İstenilen Güç için Örnek Büyüklüğü Hesaplama Yaklaşımı	31
3.GEREÇ ve YÖNTEM	33
4.BULGULAR	36

4.1.Simülasyon Çalışmasına ait Sonuçlar	36
4.1.1.İstenilen Doğruluğa Dayalı Örnek Büyüklüğü için Simülasyon Sonuçları	36
4.1.1.1.Paralel ölçüm varsayımı ile elde edilen simülasyon sonuçları	36
4.1.1.2.Paralel dışı ölçüm varsayımı ile elde edilen simülasyon sonuçları	44
4.1.2.İstenilen Güce Dayalı Örnek Büyüklüğü için Simülasyon Sonuçları	49
4.2.Gerçek Veri Setine ait Bulgular	53
4.2.1.İstenilen Doğruluğa Dayalı Örnek Büyüklüğü Değerlendirilmesi	53
4.2.1.1. Paralel Ölçüm Durumunda Örnek Büyüklüğünün Değerlendirilmesi (ODI)	53
4.2.1.2. Paralel Dışı Ölçüm Durumunda Örnek Büyüklüğünün Değerlendirilmesi (SF-36)	57
4.2.1.2.1. Örnek Büyüklüğü Çalışması Sonuçları	59
4.2.1.2.2. Güven Aralığı Çalışması Sonuçları	65
4.2.2.İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirilmesi	67
4.2.2.1.ODI Verisinde İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirmesi	67
4.2.2.1.SF-36 Verisinde İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirmesi	69
4.2.2.1.1.SF-36 Alt Ölçekleri ile Güç Değerlendirmesi	69
4.2.2.1.2.Bir Bütün Olarak SF-36 ile Güç Değerlendirmesi	72
5.TARTIŞMA	75
6.SONUÇ ve ÖNERİLER	80
7.KAYNAKLAR	82
EKLER	85
R kodları	85
Simülasyon için Kodlar	85
Gerçek Veri için Kodlar	88
ÖZGEÇMİŞ	97

ÇİZELGELER DİZİNİ

Çizelge 4.1.1. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.05 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	37
Çizelge 4.1.2. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.1 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	39
Çizelge 4.1.3. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.15 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	41
Çizelge 4.1.4. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.05 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	45
Çizelge 4.1.5. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.1 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	46
Çizelge 4.1.6. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.15 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar	47
Çizelge 4.1.7. $H_0: \alpha = h$ hipotezini test etmek için 0.80, 0.85 ve 0.90 güç değerlerine, farklı madde sayılarına ve planlanan Cronbach alfa değer aralıklarına dayalı örnek büyüklüğüne ait tanımlayıcılar	50
Çizelge 4.2.1. ODI'nin, elemanları %95'lik güven aralığından oluşan varyans-kovaryans matrisi	54
Çizelge 4.2.2. ODI'den rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler	55
Çizelge 4.2.3. SF-36 alt ölçeklerine ait korelasyon analizi	57
Çizelge 4.2.4. FF'in, elemanları %95'lik güven aralığından oluşan varyans-kovaryans matrisi	58
Çizelge 4.2.5. FF ölçeğinden rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler	59
Çizelge 4.2.6. SF-36 alt ölçeklerinin (FRG, ERG, V, VA, GSA, Sİ, RS) rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler	62
Çizelge 4.2.7. SF-36 ölçeğinin rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler	64
Çizelge 4.2.8. SF-36 ölçeğinin FF-GSA-RS alt boyutlarına paralel dışı simülasyon çalışmasında denk gelen örnek büyüklüğü sınıfı	66

Çizelge 4.2.9.ODI ölçeğinin rastgele örnekleme ile elde edilen güç değeri için tanımlayıcı istatistikler	67
Çizelge 4.2.10.ODI'nin, istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı	68
Çizelge 4.2.11.SF-36 alt ölçeklerinin $H_0: \alpha = h$ hipotezindeki h değerleri	69
Çizelge 4.2.12.SF-36 alt ölçeklerinin rastgele örnekleme ile elde edilen güç değerleri için tanımlayıcı istatistikler	70
Çizelge 4.2.13.SF-36 alt ölçeklerinden FF, GSA ve RS'in istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı	72
Çizelge 4.2.14.SF-36 ölçeğinin rastgele örnekleme ile elde edilen güç değeri için tanımlayıcı istatistikler	73
Çizelge 4.2.15.SF-36 ölçeğinin, istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı	74

ŞEKİLLER DİZİNİ

Şekil 2.3.1.1. Geçerlik Çeşitleri	10
Şekil 2.3.2.1. Güvenirliğin Temel Bileşenleri ve Belirlenmesinde Kullanılan Yöntemler	16
Şekil 4.1.1. Paralel ölçüm durumunda, istenilen güven aralığı genişliği 0.05 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	38
Şekil 4.1.2. Paralel ölçüm durumunda istenilen güven aralığı genişliği 0.1 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	40
Şekil 4.1.3. Paralel ölçüm durumunda istenilen güven aralığı genişliği 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	42
Şekil 4.1.4. Paralel ölçüm durumda istenilen güven aralığı genişliği 0.05, 0.1 ve 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	43
Şekil 4.1.5. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliği 0.05, 0.1 ve 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	48
Şekil 4.1.6. Planlanan farklı Cronbach alfa değer aralıklarında $H_0: \alpha = h$ hipotezini test etmek için gerekli örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve güce bağlı dağılımı	51
Şekil 4.1.7. İstenilen güç değeri 0.8, 0.85 ve 0.9 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı	52
Şekil 4.2.1. ODI'de Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	55
Şekil 4.2.2 FF ölçeğinde Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	60
Şekil 4.2.3. SF-36 alt ölçeklerinde Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	63

Şekil 4.2.4. SF-36 alt ölçeklerinin Cronbach alfa değerleri ortalaması ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	64
Şekil 4.2.5. SF-36 ölçeğinin Cronbach alfa değerleri ortalaması ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	65
Şekil 4.2.6 ODI’de güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	68
Şekil 4.2.7. SF-36 alt ölçeklerinde güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	71
Şekil 4.2.8. SF-36 ölçeğinde güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı	73



SİMGELER ve KISALTMALAR DİZİNİ

MTMM	Multitrait-Multimetod
CVR	Content Validity Ratio - İçerik geçerliği oranı
CVI	Index of Content Validity - İçerik geçerliği katsayısı
KR₂₀	Kuder Richardson 20 Katsayısı
ODI	The Oswestry Disability Index – Oswestry Bel Ağrısı Ölçeği
SF-36	Short From-36
FF	SF-36 fiziksel fonksiyon alt boyutu
FRG	SF-36 fiziksel rol güçlüğü alt boyutu
ERG	SF-36 emosyonel rol güçlüğü alt boyutu
V	SF-36 vitalite alt boyutu
VA	SF-36 vücut ağrısı alt boyutu
GSA	SF-36 genel sağlık algısı alt boyutu
Sİ	SF-36 sosyal işlevsellik alt boyutu
RS	SF-36 ruhsal sağlık alt boyutu

ÖZET

Güvenirlik Çalışmalarında Örnek Büyüklüğü Hesaplama Yaklaşımları

Sağlık alanında kullanılan ölçeklerin standart ölçüm aracı olabilmesi için “geçerlik” ve “güvenirlik” temel özelliklerine sahip olması gerekir. Ölçek güvenilirliğini belirlemek adına birçok yöntem mevcuttur ve Cronbach alfa güvenilirlik katsayısı sağlık bilimlerinde en çok kullanılan güvenilirlik ölçütlerinden biridir. Güvenirlik analizlerinde önemli bir konu, “bir ölçekte hesaplanacak Cronbach alfa katsayısını doğru olarak belirleyebilmek için gerekli olan örnek büyüklüğü nedir” sorusunu yanıtlamaktır. Cronbach alfa güvenilirlik katsayısının popülasyon değerinin tahmininde genel olarak yararlanılan güven aralığı yöntemi, çoklu ölçümlerin eşit varyansa ve kovaryansa sahip olması (paralel ölçüm) gibi sınırlayıcı varsayımları barındırmaktadır. Bonett, 2015 yılında eşit varyanslara veya eşit kovaryanslara gereksinim duyulmayan (paralel dışı ölçüm) bir güven aralığı yaklaşımı kullanılarak istenilen doğruluk ve hipotez testlerinde istenilen güce ulaşmak adına örnek büyüklüğü hesaplamaları önermiştir.

Bu tez ile ölçeklerde güvenilirlik çalışmalarının planlanması aşamasında bu iki örnek büyüklüğü yaklaşımının, farklı Cronbach alfa katsayılarına ve ölçeklerin içerdiği farklı madde sayılarına bağlı olarak değişimlerini araştırmak amaçlanmıştır.

Bu amaçla yapılan simülasyon çalışmalarında öncelikle istenilen doğruluğa dayalı örnek büyüklüğü yaklaşımı sırasıyla; paralel ölçümler baz alınarak farklı madde sayıları, planlanan farklı Cronbach alfa seviyeleri ve farklı güven aralığı genişliklerine göre, paralel dışı ölçümlerde ise bunlara ek olarak varyans-kovaryans matrisi yapılarına göre incelenmiştir. Daha sonra istenilen güce dayalı örnek büyüklüğü, H_0 hipotezi altında farklı güç değerleri ele alınarak değerlendirilmiştir. ODI ve SF-36 gerçek veri setleriyle de paralel ve paralel dışı ölçüm durumlarına ait simülasyon çalışmaları desteklenmiştir.

Önerilen güven aralığı ve örnek büyüklüğü yaklaşımlarına dayanarak, yapılan çalışmalarda; madde sayısı, Cronbach alfa değer aralığı ve güven aralığı genişliği arttıkça örnek büyüklüğünün azaldığı, istenilen güç arttıkça ise örnek büyüklüğünün arttığına dair bulgular elde edilmiştir. Sonuç olarak, istenilen doğruluğa dayalı örnek büyüklüğü yaklaşımında ölçüm durumlarından hangisi olduğuna bakılmaksızın, ölçekte bulunan madde sayısı arttıkça örnek büyüklüğü belirli bir seviyeye kadar azalmakta ancak madde sayısı 50 olduğunda sabitlenmektedir. Madde sayısı 50 olduğunda meydana gelen örnek büyüklüğündeki bu sabitlenme, hipotez testine bağlı istenilen güç için örnek büyüklüğü yaklaşımında da görülmektedir.

Anahtar sözcükler: güvenilirlik, Cronbach alfa, örnek büyüklüğü, güven aralığı genişliği, güç

ABSTRACT

Sample Size Calculation Approaches in Reliability Studies

A standard scale used in the fields of health must have two basic properties: reliability and validity. There are many methods to determine the reliability of scale and the Cronbach alfa reliability coefficient is one of the most widely used reliability measurement in the health sciences. The important point in the analysis is deciding the sample size needed for correctly identifying Cronbach alfa coefficient of the scale. The classical confidence interval method for the population value of Cronbach's alfa reliability coefficient has a restrictive assumption that the multiple measurements have equal variances and equal covariances (i.e. parallel measurement). In 2015, Bonett proposed the use of a confidence interval approach that does not require equal variances or/and equal covariances (nonparallel measures) for calculating sample size to achieve the desired precision and to reach the desired level of power in hypothesis tests.

This study aims to investigate the changes in sample size created by these two sample size approaches in general, depending on the different Cronbach alfa coefficients and the number of items included in the scales.

Simulation studies are performed for this purpose. Firstly, for the sample size depending on the desired precision based on the parallel measurements; different item counts, varying planned Cronbach alfa levels and confidence interval widths were used. In the case of non-parallel measures, the variance-covariance matrix is examined in addition to those mentioned. Later, the sample size depending on the desired power was evaluated by taking different power values under the H_0 hypothesis. Simulation studies of parallel and non-parallel measurement situations were supported by ODI and SF-36 real datasets.

Based on the proposed confidence interval and sample size approaches; findings indicate that sample size decreases as the number of items, Cronbach alpha value range and confidence interval width increase. Increasing the required power causes a increase in the sample size. As a result, in the sample size approach based on the desired precision, regardless of the model type (parallel measure or not), as the number of items in the scale increases, the sample size decreases to a certain level, but is fixed when the item number is 50. This fixation of sample size at 50 items was also observed in the hypothesis test method (desired power) for sample size determination.

Keywords: reliability, Cronbach alpha, sample size, confidence interval width, power

1. GİRİŞ

Bilimde ölçme önemli bir yere sahiptir. Stevens (1951) tarafından ölçme, “olaylara ya da nesnelere kurallara göre atama yapma işlemi” olarak tanımlanmaktadır. Bir başka deyişle ölçme, soyut kavramları gözleme dayalı göstergelere bağlama sürecidir¹. Soyut kavramları ölçmek, başta sağlık (özellikle psikoloji ve fizik tedavi) ve davranış bilim dalları olmak üzere bütün bilim dallarında giderek daha fazla önem kazanmaktadır, ancak soyut kavramların ölçülmesi istendiği sosyal bilimler ve sağlık bilimlerinde ölçme işlemi fen bilimlerine göre daha zordur.

Ölçmenin tanımı ile birlikte, deneysel bir göstergenin, verilen teorik kavramı temsil etme derecesinin nasıl belirleneceği sorusu ortaya çıkar. Burada devreye ölçüm araçları girer¹. Ölçüm araçlarının bir türü olan ölçek, sayısal değerler elde etmek için kullanılan belirli bir birimde derecelenmiş araç ya da gereçler olarak adlandırılır. Pekçok alanda fiziksel yöntemlere dayalı ölçme araçları kullanılır. Örneğin, somut bir nitelik olan ağırlığı doğrudan kantar ile ölçmek mümkündür. Bununla birlikte kişilerin davranış, kişilik özellikleri, vb. soyut özelliklerinin doğrudan ölçülmesi için fiziksel araçlar bulunmadığından farklı ölçme araçları geliştirilerek dolaylı yoldan ölçüm elde edilir. Doğrudan ölçülemeyen özellikler için, ölçülmek istenen amaca yönelik sorulan maddelerden oluşan ölçekler geliştirilir².

Sağlık alanında kullanılan ölçeklerin, standart bir ölçüm aracı olması için “güvenirlik” ve “geçerlik” ana özelliklerine sahip olması istenir³. Araştırmacıların, bir veya daha fazla potansiyel ilgi ölçeği bulunan bir kavramın mevcut ölçeklerinin hangisini kullanacağını seçme konusunda karar verme veya yeni bir ölçek geliştirme süreci, bu ana özelliklerin değerlendirilmesiyle tamamlanır⁴. Bir ölçeğin güvenilirliği, bir ölçüm sürecinde, ölçüm işleminin tekrarlanması durumunda benzer sonuçların elde edilmesi iken geçerliği, ölçeğin ölçmeyi amaçladığı özelliği ölçme derecesi olarak tanımlanır. Bu iki kavram arasında bir ilişki söz konusudur. Bir ölçek eğer geçerli ise güvenilir de olmalıdır ancak tersi her zaman doğru değildir⁵.

Araştırmacılar, elde edilen bir ölçek puanını değerlendirdiğinde, puanın, kişinin gerçek puanı için iyi bir tahminde bulunmasını ister. Ölçek puanının güvenilirliği ne kadar yüksek olursa, elde edilen skorun gerçek skora o kadar yakın olması gerekir. Klinik açıdan bakıldığında, popülasyon güvenirliliğiyle ilgili kesin olmayan bir tahmin klinik kararları önemli ölçüde etkileyebilir. Geçerlik ve güvenilirlik, sağlam ve doğru sonuçlar elde etmenin yanı sıra bulguların daha geniş bir kitleye genelleştirilmesine ve sonuç olarak insanların yaşam biçimlerini iyileştirmek için araştırma sonuçlarının değerlendirilebilir ve uygulanabilir olmasına yardımcı olur⁶.

Güvenirlik çalışmalarında, ölçeklerin güvenilirliğini belirlemek adına birçok yöntem mevcuttur. Burada önemli olan nokta, bu güvenilirlik çalışmalarının planlama aşamasında, “bir ölçeğin güvenilirlik katsayısını doğru olarak belirleyebilmek için ne kadar örnek büyüklüğü almamız gerektiği” sorusuna cevap bulmaktır. Güvenirlik analizlerinin çalışma tasarımıyla örnek büyüklüğü belirlenmesi dikkate alınması gereken en önemli konulardan biridir. Örnek büyüklüğü için farklı teorik ve bulgusal yaklaşımlar kullanılmaktadır. Uygun örnek büyüklüğü önemlidir, zira bilindiği üzere küçük örnek büyüklüğü çalışmada düşük bir güce, büyük örnek büyüklüğü ise kaynak israfına yol açar⁷.

Cronbach alfa güvenilirlik katsayısı, sağlık bilimlerinde en çok kullanılan güvenilirlik ölçütlerinden biridir. Ölçeklerin güvenilirlik çalışmalarında, istenilen doğruluk (desired precision) ve istenilen güç (desired power) için olmak üzere Cronbach alfa güvenilirlik katsayısına bağlı iki farklı örnek büyüklüğü hesaplama yaklaşımı mevcuttur⁸.

Bu tez çalışmasının amacı,

- Ölçeklerde, güvenirliliğin hesaplanmasında en çok kullanılan güvenilirlik katsayısı olan Cronbach’ın alfası ile ilgili teorik çerçeveyi ortaya koymak,
- Cronbach’ın alfasına bağlı örnek büyüklüğü yaklaşımlarını incelemek,
- Güvenirlik çalışmalarının planlanması aşamasında, örnek büyüklüğü yaklaşımlarının:
 - Farklı Cronbach alfa katsayılarına,
 - Ölçeklerin içerdiği madde sayısına,
 - Farklı güven aralığı genişliklerine,
 - Farklı güç seviyelerine bağlı olarak

değişimlerini araştırmaktır.

2. GENEL BİLGİLER

Bir konuda ölçüm aracı geliştirilmesinin sebebi, kişilerin o konudaki gerçek durumlarını en doğru şekilde tahmin etmek olduğundan, iyi bir ölçeğin bazı değerlendirme süreçlerinden geçirilerek hazırlanması gerekir⁹. Burada değerlendirme sürecinin içeriğini, geçerlik ve güvenirlik kavramları oluşturur. Veri toplama sürecinin bir parçası olarak bir ölçüm aracı kullanıldığında, kullanılan ölçüm aracının geçerliği ve güvenirliği önemlidir. Ölçüm güvenirliğini sağlamak, geçerliği sağlamaya yönelik gerekli bir ilk adımı oluşturduğundan, hem uygulamalı hem de teorik araştırmalarda tartışılmaz bir öneme sahiptir¹⁰. Güvenirlik değerlendirmesi ayrıca nicel sonuçların doğru yorumlaması için de önemlidir¹¹.

2.1. Ölçeklerde Geçerlik ve Güvenirlik

Mevcut sağlık durumu ölçeklerinin sayısı son on yılda çarpıcı bir biçimde artmıştır. Bu durum, aynı kavramı ölçen farklı ölçeklerden hangisinin kullanılacağı seçimini büyük bir zorluk haline getirmiştir. Sağlık durumu ölçeklerinin geliştirilmesi ve değerlendirilmesine yönelik çalışmaların metodolojik kalitesini belirlemek için kriterlere ihtiyaç vardır¹². Bu kriterlerden en önemlileri, geçerlik ve güvenirliktir. Ölçüm araçlarının geçerliğinin ve güvenirliğinin doğrulanması, çalışma bulgularının bütünlüğünün sağlanması için bir ön şarttır. Geçerlik ve güvenirlik değerlendirmeleri ile elde edilen standartlaştırılmış ölçüm araçları ölçme işleminin daha duyarlı yapılarak objektif ölçümler elde edilmesini sağlar^{1,13}.

Bir ölçeğin geçerliğinden bahsedilince akla gelen, ölçek maddelerinin ölçülmek istenen kavramı doğru bir şekilde yansıtip yansıtmadığının incelenmesidir. Güvenirlik ise, aynı kavramı ölçmede farklı maddelerin ne kadar tutarlı olduğunun incelenmesidir^{1,2}. Günlük hayattan örnek vermek adına, okçuluğa ilgi duyan bir birey ele alınsın. Okçuluğa merak duyan birinin ilk öğrenmesi gereken şey hedefin merkezini nasıl vuracağıdır ki bunu daha sonra tutarlı bir şekilde yapabilir hale gelsin. Bu durum bir ölçümün geçerlik ve güvenirliği ile paraleldir. Burada geçerlik, atış amacı ile temsil edilecektir(ortalama olarak atışların hedefin merkezine ne kadar yakın olduğu ile). Ölçümün tutarlılığı ise,

hedefte her nereye gelmiş olursa olsun, ardışık olarak atılan okların birbirlerine ne kadar yakın olduğu ile belirlenebilir. Bir grup atış hedefin merkezinde yer alıyorsa geçerli ve güvenilir olarak nitelendirilirken bir grup atışın hedefin merkezinden uzaklaşması testin güvenilir fakat geçerli olmadığını belirtmektedir¹⁴.

2.2. Klasik Test Teorisi

2.2.1. Klasik Test Teorisinin Tanımlanması

Klasik test teorisi, güvenilirliğin temelini oluşturur ve ölçek skorunun, yalnızca test kavramının ölçülme kabiliyeti üzerine gerçek skorunu değil aynı zamanda hata skorunu da yansıttığını gösterir. Kişinin gözlenen skoru ile gerçek skoru arasındaki fark ölçme hatalarından kaynaklanır. Hatasız bir ölçme teorik olarak mümkün olmadığından istenen, hatayı minimize ederek güvenilirliği yükseltmektir^{1,2,15,16,17}.

Klasik test teorisi modeli,

$$x_i = \tau_i + \varepsilon_i \quad (1)$$

şeklindedir^{1,15,18}. Eşitlik (1)'de,

x_i : i'nci gözlenen skor (test skoru)

τ_i : i'nci gerçek skor

ε_i : $x_i - \tau_i$ olarak tanımlanan rasgele ölçüm hatasıdır.

Bireyin test sonucunda elde edilen skoruna gözlenen skor denir ve Eşitlik (1)'de görüldüğü gibi gerçek skor ve hata skorunun toplamından oluşur. Gerçek skor ise, varsayımsal olarak ifade edilen, direk ölçülemeyen, gözlemlenememiş miktardır. Sonsuz sayıda tekrarlanmış ölçümün ortalaması gerçek skora eşit olabilir. Ancak, sonsuz sayıda tekrarlanan ölçüm elde etmek mümkün olmadığından gerçek skor varsayımsal, hiçbir zaman bilinmeyen bir niteliktir. Ayrıca, Eşitlik (1), herhangi bir gözlenen skorun rasgele hatalar sebebiyle hiçbir zaman gerçek skora eşit olmayacağını da göstermektedir^{1,4,19}.

Klasik test teorisinin genel varsayımları aşağıdaki gibidir^{1,15,17}.

- Ölçüm hatalarının beklenen değeri sıfıra eşittir.

$$E(\varepsilon_i) = 0$$

- Gerçek skorlarla hata skorları arasındaki kovaryans sıfıra eşittir.

$$\text{cov}(\tau_i, \varepsilon_i) = 0$$

- İki ölçümün hata skoru arasındaki korelasyon sıfıra eşittir.

$$\rho(\varepsilon_i, \varepsilon_j) = 0$$

- i 'nci ölçümün hata skoru ile $(i + 1)$ 'nci ölçümün gerçek skoru arasındaki korelasyon sıfıra eşittir.

$$\rho(\varepsilon_i, \tau_{i+1}) = 0$$

Eşitlik (1), aynı zamanda varyanslar için de söz konusudur. Gözlenen skorun varyansı, σ_x^2 ; gerçek skorun varyansı olan σ_τ^2 ile hata skorları varyansı olan σ_ε^2 'nin toplamına eşittir^{1,5}.

$$\sigma_x^2 = \sigma_\tau^2 + \sigma_\varepsilon^2 \quad (2)$$

Güvenirlilik, Eşitlik (2) yardımıyla; gerçek skor varyansının gözlenen skor varyansına oranı olarak elde edilir²⁰.

$$\rho_x = \frac{\sigma_\tau^2}{\sigma_x^2} = \frac{\sigma_x^2 - \sigma_\varepsilon^2}{\sigma_x^2} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_x^2} \quad (3)$$

Ölçüm güvenirliğinin, 0 ile 1 arasında değer aldığı eşitlik (3) yardımıyla görülür^{5,21}. Hata varyansı arttıkça, ki bu gözlenen skor varyansını da etkiler, güvenirlik sıfıra yaklaşır. Hata varyansının azalması durumunda ise, güvenirlik maksimum seviyeye yaklaşır¹¹. Klasik test teorisinde dikkat edilmesi gereken en önemli nokta, bireylerin gerçek skoru bilinmediğinden, σ_τ^2 'nin hesaplanmasının pratikte pek mümkün olmadığıdır^{1,5}.

2.2.2. Klasik Test Teorisi Modelleri

Çıkış noktası Eşitlik (1) olan, gerçek skorlar için farklı indeksler kullanılmasına rağmen tek bir gerçek skora(τ) sahip ve ölçüm hataları ε_i ve ε_j 'nin varyanslarının sırasıyla $\sigma_{\varepsilon_i}^2$ ve $\sigma_{\varepsilon_j}^2$ olduğu i ve j ölçek maddeleri aşağıdaki gibi tanımlanır^{18,21}.

$$x_i = \beta_i \tau_i + \varepsilon_i \quad (4)$$

$$x_j = \beta_j \tau_j + \varepsilon_j$$

Bu varsayımlar altında 3 farklı ölçüm modeli mevcuttur ve aşağıdaki gibidir.

- **Paralel model:** Eşitlik (4)'de $\beta_i = \beta_j = 1$ ve $\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_j}^2$ ise 'paralel model' söz konusudur. Paralel model, gerçek skoru tanımlamada kullanılacak en kısıtlayıcı ölçüm modelidir. Paralel model, tüm ölçek maddelerinin tek bir değişkeni (tek boyutluluk) ölçmesini gerektirirken, aynı zamanda tüm maddelerin birbirine eşdeğer olduğunu varsayar. Bu model söz konusu olduğunda, tüm maddeler aynı sonuç değişkenini, aynı ölçekte, aynı derecede hassasiyetle ve aynı miktarda hata ile ölçmelidir. Tüm maddelerin gerçek skorlarının birbirine eşit olduğu varsayımına ek olarak, maddelerin hata skorları da benzer şekilde eşittir²².

- **Tau-eşdeğer(tau-equivalent) ve esasen tau-eşdeğer(essentially tau-equivalent) model:** Eşitlik (4)'de $\beta_i = \beta_j = 1$ ve $\sigma_{\varepsilon_i}^2 \neq \sigma_{\varepsilon_j}^2$ ise 'tau-eşdeğer model' söz konusudur.

Tau-eşdeğer model, daha az kısıtlayıcı olmasıyla birlikte paralel modelle benzerdir. Maddelerin hata varyanslarının birbirinden farklı olduğu durumlarda kullanılır. Bu, maddelerin aynı sonuç değişkenini aynı ölçekte, aynı derecede hassasiyetle, ancak farklı hata miktarlarıyla ölçtüğü anlamına gelir.

Esasen tau-eşdeğer model, her maddenin aynı yanıt değişkenini aynı ölçekte, ancak muhtemelen farklı derecelerde hassasiyetle ölçtüğünü varsaymaktadır. Tau-eşdeğer modelinde olduğu gibi, esasen tau-eşdeğer modeli de hata

varyanslarının birbirinden farklı olması durumunda kullanılır. Tau-eşdeğer model, maddelerin gerçek skorlarının eşit olduğunu varsaymakta iken, esasen tau-eşdeğer model, her maddenin gerçek skorunun, sabit bir fark ile toplamsal olarak farklı olduğunu varsayar²².

- **Konjenerik(congeneric) model:** Eşitlik (4)'de $\beta_i \neq \beta_j = 1$ ve $\sigma_{\varepsilon_i}^2 \neq \sigma_{\varepsilon_j}^2$ ise 'konjenerik model' söz konusudur. Konjenerik model, güvenirlik tahmini için en az kısıtlayıcı, en genel modeldir. Konjenerik model, her bir maddenin aynı sonuç değişkenini, farklı ölçeklerle, farklı derecelerde hassasiyetle ve farklı hata miktarlarıyla ölçtüğünü varsayar. Esasen tau-eşdeğer model, madde skorlarının sadece bir sabit ile farklı olmasına izin verdiği halde, konjenerik model, maddelerin gerçek skorları arasında doğrusal bir ilişki olduğunu varsayar ve her bir madde gerçek skor çifti arasında hem toplamsal hem de çarpımsal bir sabite izin verir²².

Paralel model, psikometride temel ölçüm modelidir. tau-eşdeğer model ve konjenerik model ise paralel modelin biraz daha genel ifade edilmiş biçimidir^{18,21}.

2.2.3. Geçerlik ve Güvenirlik Çalışmalarında Hatalar

Geçerlik ve güvenirlik çalışmalarında sistematik ve rasgele hatalar ortaya çıkabilir, hatta Judd, Smith ve Kidder(1991) bu hataları ayrı ayrı dikkate alan bir model (Eşitlik 5) de önermişlerdir⁴. Rasgele hatalar, nedeni belli olmayan, ölçüm sonuçları üzerinde her seferinde aynı etkiyi göstermeyen ve saptanması zor olan hatalardır. Sistematik hatalar ise sabit hata ve oransal hata olmak üzere iki çeşittir. Sabit hata, bir ölçümden diğer ölçüme değişmeyen hata türüdür. Ölçümlerdeki hata miktarı aynı ve sebebi bellidir. Bir kantarın bireyin ağırlığını her seferinde 2 kilo eksik veya fazla ölçmesi sabit hata durumuna örnek olarak verilebilir. Oransal hata ise, ölçülen değişkenin büyüklüğüne göre oranı değişen yani belirli bir kurala dayanan hatadır. Hata kaynağı bellidir ve düzeltilebilir. Bir öğretmen matematik sınavında yazısı güzel olanların sınav puanına, sınav puanının %5 fazlasını da ekliyorsa bu durum oransal hataya bir örnektir⁵.

Klasik Test Teorisinin genel eşitliği (1) hata türleri ile genişletilerek Eşitlik(5)'teki model elde edilebilir⁴.

$$skor_{gözlenn} = skor_{İY} + \varepsilon_{SH} + \varepsilon_{RH} \quad (5)$$

Eşitlik (5)'te,

$skor_{İY}$: İlgilenilen yapıya ait skor (construct of interest)

ε_{SH} : Sistematik hata

ε_{RH} : Rasgele hata

şeklinde ifade edilir. Eşitlik (5)'teki $skor_{İY} + \varepsilon_{SH}$ ile Eşitlik (1)'deki τ_i aynıdır.

Gerçek skoru, Eşitlik (5)'teki gibi ilgilenilen yapı skoru ve sistematik hataların toplamı olarak yazmanın iki avantajı vardır. Birincisi, geçerlik ve güvenirlik arasındaki ilişkinin rahatlıkla gösterilmesi, ikincisi ise farklı hata türlerinin geçerlik ve güvenirliği nasıl etkilediğinin incelenmesidir.

Eşitlik (5) dikkate alınırsa,

$$Güvenirlik = \frac{\sigma_{İY}^2 + \sigma_{SH}^2}{\sigma_{gözlenn}^2}$$

$$Geçerlik = \frac{\sigma_{İY}^2}{\sigma_{gözlenn}^2}$$

şeklinde elde edilir. Bu iki eşitlikten de anlaşıldığı gibi, rasgele hata hem güvenirliği hem de geçerliği etkilemekteyken sistematik hata sadece geçerliği etkilemektedir^{5,23}. Yani, güvenirlik; rasgele hata miktarı ile ters ilişkiliyken, geçerlik; ölçüm sürecinde var olan rasgele olmayan hata miktarına da bağlıdır¹.

2.3. Geçerlik

İyi klinik, araştırma ve eğitim uygulamaları için sağlam değerlendirme yöntemleri gerektir. Bu sebeple, semptom ölçekleri, eğitim testleri ve gözlemci puanları gibi ölçüm araçlarından alınan puanların geçerliği değerlendirilmelidir. Geçerlik, "kanıtların ve

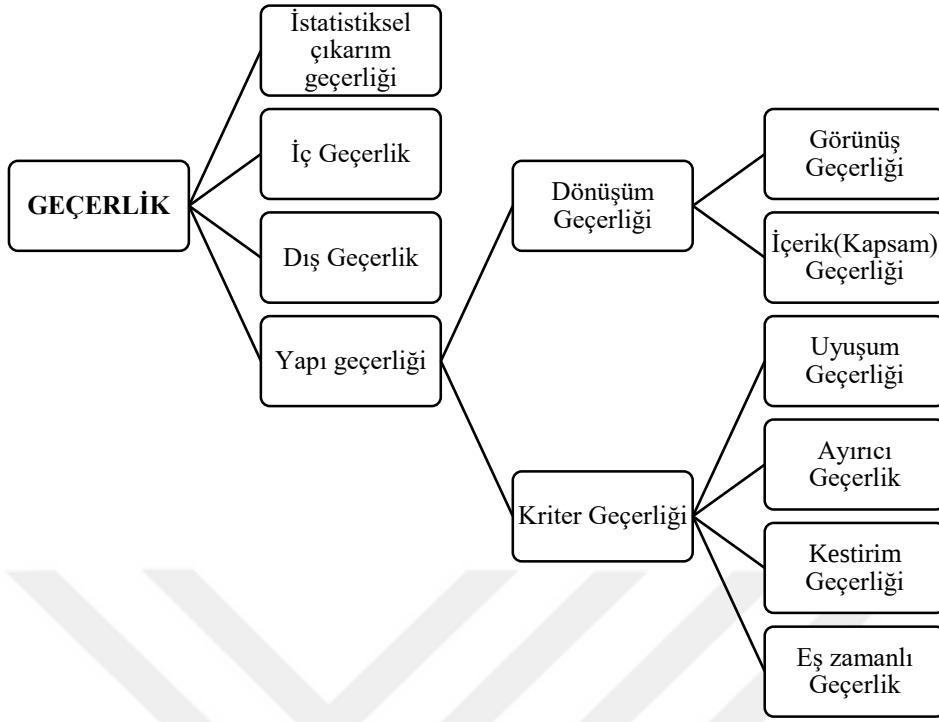
teorinin, önerilen test kullanımları ile verilen test puanlarının yorumlarını destekleme derecesini" ifade eder²⁴.

Geçerlik, ölçeğin ölçmeyi amaçladığı özelliği ölçme derecesidir^{1,5}. Yani, verilerin ölçülmek istenilen özelliğin niteliğini tam olarak yansıtmaması olarak tanımlanır². Ölçme aracı ile elde edilen verilerin amaca yönelik olarak kullanılabilir olması o ölçme aracının geçerli olduğunun göstergesidir²⁵. Bununla birlikte, bu tanım sınırlıdır, çünkü bazen testin orijinal amacı dışına çıkan çıkarımlar yapılabilir. Bir kitabı kapı çarpmasını engellemek amacıyla kapı aralığında durdurucu olarak kullanmak gibi, bir ölçümün başlangıçta tasarlananlardan başka kullanım alanları olabilir¹⁴.

Geçerlik tamamen güvenilirliği esas almaktadır, güvenilirlik ise geçerlik için gereklidir ancak yeter koşul değildir. Yani, güvenilir bir ölçek geçerli olmayabilir fakat geçerli bir ölçek muhakkak güvenilir olmalıdır^{1,5,25}.

2.3.1. Geçerlik Çeşitleri

Bir ölçek ile, ölçülmek istenilen özellik doğru ve diğer özelliklere karıştırılmadan ölçülüyor ise geçerlikten söz edilebilir. Geçerli bir ölçek, belirli amaçları yerine getirmede başarılı olan ölçektir^{1,5}. Bu kapsamda, literatürde geçerlik tiplerine ilişkin yaygın olarak kullanılan farklı sınıflandırmalar mevcuttur.



Şekil 2.3.1.1. Geçerlik Çeşitleri

Temel olarak 4 tip geçerlik söz konusudur.

- I. **İstatistiksel Çıkarım Geçerliği (Statistical Conclusion Validity):** İstatistiksel çıkarım geçerliği test edilen ilişki ile ilgilidir. Aynı zamanda, belirli alfa değerinde kovaryasyonun (covariation) tahmininin ve elde edilen varyansın makul olup olmadığı hakkındaki çıkarımı ifade eder. Düşük istatistiksel güç, testlerde varsayımların bozulması, ölçüm güvenilirliğinin düşük olması, katılımcıların heterojenliği gibi istatistiksel çıkarım geçerliğini etkileyen bazı önemli tehditler bulunmaktadır²³.
- II. **İç Geçerlik (Internal Validity):** İç geçerlik, ilginin kaynağının kapsamlı olarak ölçekteki maddelerle temsil edilme derecesini inceler²⁴.
- III. **Dış Geçerlik (External Validity):** Dış geçerlik, ölçekleri farklı kişilere, normlara ve zamanlara genelleme imkânı verir. Yani, ölçeklerin uygulanmasıyla elde edilen araştırma sonuçlarının ulusal veya uluslararası toplumlara genellenebilirliğidir^{23,25}.

IV. Yapı Geçerliği (Construct Validity): Yapı geçerliği, doğrudan ölçülemeyen bir kavramın, fikrin ya da davranışın, doğrudan ölçülebilen bir işleyişe dönüştürülme derecesini ifade eder. Yani, ölçülemez soyut kavramların ölçülebilir hale getirildiği durumdaki iyiliğini temsil eder^{5,23}. Aynı zamanda, ölçülmek istenen yapının tam ve yansız olarak ölçülebilmesi adına ölçeğin kurgulanması olarak tanımlanabilir²⁵.

Yapı geçerliği, bir ölçüm aracının ölçmeyi amaçladığı yapıyı ölçme derecesidir. Ölçüm aracının maddeleri, tanımlanmış teori ve kavramlarıyla ilişkilirse yapı geçerliği desteklenir. Örneğin, anksiyeteyi ölçmek için tasarlanmış bir ölçüm aracındaki tüm maddeler teorik ve yapısal olarak sadece anksiyete ile ilgili kavramları ölçme yeteneğine sahipse yapısal olarak geçerlidir. Bununla birlikte, ölçüm aracı aynı zamanda stres veya depresyon gibi yakından ilgili kavramları ölçme yeteneğine sahipse, anksiyete ölçümü olarak yeterli yapı geçerliği olmayabilir.

Faktör analizi veya Multitrait-Multimetod (MT-MM) yaklaşımı bir ölçüm aracının yapı geçerliğini değerlendirmenin çeşitli yollarıdır¹³.

Yapı geçerliğinin sağlanması, 6 çeşit geçerlik tipinin varlığını gerektirir. Bu geçerlik tipleri, dönüşüm geçerliği ve kriter geçerliği olmak üzere iki ana başlıkta sınıflandırılmıştır²³.

IV.1. Dönüşüm Geçerliği (Translation Validity): Dönüşüm geçerliği, ölçülemez soyut kavramlar bir işleyişe dönüştürüldüğünde yapının doğru anlamının yansıtıp yansıtılmadığı ile ilgilidir. Görünüş geçerliği ve içerik geçerliğini kullanarak yapıların işleyişe doğru bir şekilde dönüştürülme derecesini temsil eder²³.

IV.1.1. Görünüş geçerliği (Face Validity): Görünüş geçerliği, ölçekte bulunan maddelerin ilişkili olup olmadığını değerlendirmekte ve istatistiksel prosedürler kullanılarak test edilmemektedir²⁶.

Görünüş geçerliği, genellikle, ampirik test olmaması durumunda geçerliğin görünümünü tanımlamak için kullanılır. Araştırmacılar, dilbilgisi, sözdizimi, uygunluk ve ölçüm aracının gözden geçirilmesi için uzmanlar arar veya kişiler

eğitirler ve değerlendirmede sonuçların mantıksal olarak aktığı görüldüğünde görünüş geçerliği için onay verirler. Bu, bir aracın dışa dönük görünümüne dayalı olarak hızını kestirmeye veya kaldırımdan bakıldığında binanın yapısal bütünlüğünü tahmin etmeye benzer²⁴.

Görünüş geçerliği ile ilgili, görünüşe dayalı yargıların yanılabilirliği, geliştiriciler ve kullanıcılar arasında farklı algılamalar olma olasılığı ve görünüşten temel amacın çıkarılmasının verimsiz olabileceği gibi kaygılar belirtilir. Fakat yapı geçerliğini desteklemenin en kolay yolu Görünüş geçerliğinin varlığı olduğundan, görünüş geçerliği sıklıkla literatürde yer almaktadır. Görünüş geçerliği subjektif bir değerlendirmedir, bu yüzden genellikle yapı geçerliğinin zayıf bir formu olarak görülür^{13,23}.

IV.1.2. İçerik (Kapsam) Geçerliği(Content Validity): Bollen (1989), içerik geçerliğini, kavramın etki alanının netleştirildiği ve ölçülerin etki alanını tam olarak temsil edip etmediğini değerlendiren niteliksel bir geçerlik türü olarak tanımlamıştır²³.

Ölçüm aracındaki maddeler çalışmadaki niteliği tam anlamıyla örneklerse, içerik geçerliği varlığından bahsedilir. Ölçek maddelerinin bir havuzunu oluşturmak için bir araştırmacı önce literatür taraması yapar, uzman görüşlerine başvurur ve popülasyon örneklemesi yaparak veya nitel araştırma yoluyla ilgilenilen yapıyı ve boyutlarını tanımlar. İçerik uzmanlarından oluşan bir grup daha sonra potansiyel ölçek maddelerini gözden geçirir ve eğer bunlar yapının uygun göstergeleri ise ölçekte bulunmaları için onay verirler. Ölçüm aracı geliştirmenin başlangıç aşamaları, mümkün olan en geniş potansiyel madde havuzunu içermelidir ve içerik incelemelerine dayanarak madde sayısı azaltılmalıdır. İlgili yapısı çok soyutlandığında içerik analizinin değerlendirilmesi zor olabilir, ancak süreci nicelleştirmek için yöntemler önerilmiştir. İçerik geçerliğini belirlemek adına içerik geçerliği oranı (Content Validity Ratio) ve içerik geçerliği katsayısı (Index of Content Validity) önerilir^{26,27,28}.

- **İçerik geçerliği oranı (CVR):** Lawshe (1975), uzmanların her maddeyi 3 puanlık bir ölçekte değerlendirdiği bir yöntem önermiştir. CVR, 0

(anlaşma yok) ile 2 (mükemmel anlaşma) arasında değişen puanlar ile hesaplanabilir. Her bir ölçek maddesi için CVR aşağıda verilmiştir.

$$CVR = \frac{n_e - N/2}{N/2}$$

n_e : Madde gerekliliğini ‘zorunlu’ olarak seçen uzman sayısı

N : toplam uzman sayısı

Örneğin, 10 uzmandan 8’i madde gerekliliğini ‘zorunlu’ olarak seçmişse $CVR = \left(\frac{8-5}{5}\right) = 0.60$ olmalıdır. Dikkat edilmelidir ki, eğer yarıdan az sayıda uzman ilgilenilen maddenin zorunlu olduğunu bildirirse CVR negatif değer alacak, eğer $N/2$ sayıda uzman ilgilenilen maddenin zorunluluğunu bildirirse CVR sıfır değerini alacak ve eğer bütün uzmanlar ilgili maddenin zorunlu olduğunu bildirirse CVR ‘1’ değerini alacaktır. Maddenin eklenmesi için minimum CVR skoru tablosu, 0.05 önem düzeyinde tek-kuyruklu bir teste dayanılarak geliştirilmiştir. İçerik Geçerliği İndeksi (CVI) olarak adlandırılan tüm ölçüm aracının skoru, korunan maddelerin tümü için ortalama CVR değerinin belirlenmesiyle hesaplanabilir^{26,27}.

- **İçerik Geçerliği İndeksi(CVI):** Lynn (1986), içerik geçerliğini belirlemek için iki adımlı bir yöntem önermiştir. Ölçek geliştirme aşamasında, maddeler bireysel olarak içerik uzmanları tarafından değerlendirilir. Maddelerin korunması ya da çıkartılması gerekip gerekmediğini belirlemek için, 1 = ilgili olmayan 4 = çok ilgili olmak üzere 4 puanlık bir ölçek kullanılır. İkinci aşamada, yargılama aşamasında, kalan maddelerle CVI hesaplanır. Her bir ölçek maddesi için CVI, ilgi derecesi konusunda ‘3’ veya ‘4’ puanlaması yapan uzman sayısının toplam uzman sayısına oranı ile hesaplanır. Örneğin, 6 uzmandan 4’ü ilgi derecesini ‘3’ veya ‘4’ olarak puanlasın. $CVI = \frac{4}{6} = 0.67$ olarak bulunur. 0.05 önem düzeyinde hazırlanmış tablo değerlerine bakıldığında 6

uzmanın bulunduğu bir heyette CVI değerinin 0.83 olması durumunda içerik geçerliğinden bahsedilebilir. Bu durumda bu maddenin atılmasına karar verilir.

Bütün ölçek için CVI, CVI değerlerine göre varlığı korunan maddelerin başlangıçtaki madde sayısına oranı ile bulunur. Örneğin 80 maddeden 60'ı içerik geçerli bulunduysa, $CVI = \frac{60}{80} = 0.75$ olarak bulunur.

Bir maddenin varlığını koruması için yedi veya daha az uzman arasında tam bir mutabakat sağlanması gerekir. Uzmanlar, çoğu maddeyi onayladıklarında sıklıkla ortaya çıkan şişirilmiş geçerlik tahminlerinden kaçınma gereksinimi duyduklarından içerik geçerliği için güçlü bir destek oluşturmak zordur^{26,28}.

IV.2. Kriter Geçerliği (Criterion Validity): Geçerli olarak kabul edilen kriter ölçüsü ile geliştirilmekte olan ölçeğin kıyaslanmasını kapsamaktadır²⁶. Kriter geçerliği, belirli bir ölçüm aracı puanlarının altın standartla ne ölçüde ilişkili olduğunu ifade eder. Kullanılanın gerçekten " altın standart" olduğunu ikna edici argümanlar sunuyorsa ve altın standartla arasındaki korelasyon en az 0.70 ise, kriter geçerliği için olumlu bir değerlendirme yapılabilir²⁴.

IV.2.1. Uyuşum Geçerliği (Convergent Validity): Uyuşum geçerliği, teorik olarak benzer/ilişkili olması beklenen yapılar arasındaki uygunluktur. Uyuşum geçerliğine sahip bir ölçüm aracında, maddeler arası korelasyon katsayıları yüksek olmalıdır. Bir araştırmacı, ağrı ölçmek için tasarlanmış bir ölçüm aracının, maddeler arası korelasyonları yüksekse veya ağrı ölçmek için kullanılan onaylanmış diğer ölçüm araçları ile benzer sonuçlar elde edildiyse, uyum geçerliğini doğrular. Yaşam kalitesi ölçeği ve depresyon ölçeği arasındaki ilişki de uyum geçerliği için bir örnek teşkil eder^{5,13}.

IV.2.2. Ayırıcı Geçerlik (Discriminant Validity): Uyuşum geçerliğinin aksine, ayırıcı geçerlik, ölçüm aracının teorik olarak farklı yani ölçülmek istenen özellikle ilişkisiz olan yapıları ayırt etme kabiliyetidir. Ayırıcı geçerliğe sahip bir ölçüm aracında maddeler arası korelasyon katsayıları düşük olacaktır^{5,13}.

MT-MM yaklaşımı uyuşum ve ayırıcı geçerliği belirlemek için kullanılan bir yöntemdir. İlginçtir ki, bu yaklaşım psikometrik metinlerde görülmekle birlikte, karmaşık olduğundan ve kapsamlı testler, kaynaklar gerektirdiğinden uygulandığı literatürde nadiren görülmektedir. MT-MM yöntemi iki veya daha fazla yapı, iki veya daha fazla yöntem ile ölçülürken kullanılabilir. Bu yöntemde kullanılan matris ile özellikler arasındaki korelasyon ve ilişki dereceleri gösterilir. Aynı yapıdaki farklı ölçümler birbiriyle yüksek korelasyonlu olmalıdır (uyuşum geçerliği) ve farklı yapılar ise birbirleriyle düşük korelasyon göstermelidir (ayırıcı geçerlik)¹³.

IV.2.3. Kestirim Geçerliği (Predictive validity): Kestirim geçerliği, ölçek puanlarının gelecekteki bazı kriterlerde performansı tahmin etme derecesidir. Orijinal ölçüm ve kriter değişkenleri arasındaki yüksek korelasyonlar, ölçüm aracının belirtilen kriterin geçerli bir öngörücüsü olduğu sonucunu güçlendirir¹³. Kestirim geçerliği, maddenin ya da ölçeğin gelecekte beklenen gözlemleri, olguları ne kadar iyi tahmin edeceğini ölçtüğünden, ölçeğin doğruluğunu ve kalitesini ölçmek için kullanılır da denilebilir²⁵.

IV.2.4. Eş zamanlı Geçerlik (Concurrent validity): Ölçüt ölçüyle aynı zamanda var olduğunda, eşzamanlı geçerlikten bahsedilir. Eşzamanlılık, bir testin şu anki bir olayı tahmin etme kabiliyetini ifade eder²³. Eş zamanlı kritere bağlı geçerlik, aynı anda ölçüm aracı puanlarının ilgili kriterle korele olması durumunda doğrulanır¹³.

2.4. Güvenirlik

Bir ölçeğin geçerli olması için gerekli ilk koşul güvenirliktir. Ölçeğin taşıması gereken özelliklerden biri olan güvenirlik farklı şekillerde tanımlanabilir.

- Klasik test teorisinde, güvenirlik, gerçek varyansın toplam varyansa oranı olarak tanımlanır ve bir skorun rastgele hata bulundurmama derecesini ifade eder¹⁴.
- Güvenirlik; deneyin, testin ya da herhangi bir ölçme prosedürünün tekrarlı denemelerinde aynı sonuçları ne ölçüde verdiği ile ilgilenir. Yani,

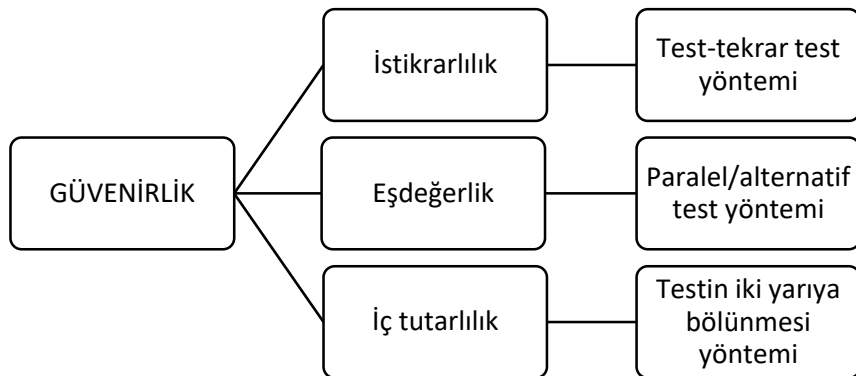
tekrarlardaki tutarlılıktır. Örneğin, bir tıbbi termometre, her kullanıldığında doğru sıcaklığı ölçen güvenilir bir ölçüm aletidir^{1,5,20}.

- Güvenirlilik, ölçeğin istikrarlılığını göstermektedir^{1,23,26}. Güvenirlilik, bir ölçüm aracının zaman içindeki tutarlı ölçüm yeteneğiyle ilgilidir^{1,20,29}. Eğer bir çalışmada kullanılan ölçüm aracı, sonuçları benzer bir metodoloji ile üretilebiliyorsa, ölçüm aracının güvenilir olduğu kabul edilir³⁰.
- Güvenirlilik, ölçümlerin kararlılığının bir göstergesidir^{1,3,5}.
- Güvenirlilik, ölçümdeki hata miktarı ile ilgilidir. Hata miktarı ne kadar azsa, ölçüm için o kadar güvenilirdir denilir¹⁴.
- Güvenirlilik, ölçeklerin uygulanması sonucunda elde edilen skorun bireyin gerçek skorunu ne ölçüde yansıttığına dair bir göstergedir⁵.

2.5. Güvenirliğin Belirlenmesinde Kullanılan Yöntemler

Güvenirlilik temelde, gerçek skor varyansının gözlenen skor varyansına bölünmesi ile elde edilir. Fakat gerçek skorların bilinmemesi, güvenirliğin bu tanımının teoride kalmasına sebep olur. Bundan dolayı, güvenirliği değerlendirmek için farklı yöntemler geliştirilmiştir⁴.

Güvenirliğin kapsamı, genel olarak bileşenlerinin içeriği ile ilgilidir. Güvenirliğin temel bileşenleri literatürde üç başlıkta toplanır. Bunlar; istikrarlılık(stability), eş değrlilik (equivalence) ve iç tutarlılıktır (internal consistency)^{2,16,20,23,26}.



Şekil 2.3.2.1. Güvenirliğin Temel Bileşenleri ve Belirlenmesinde Kullanılan Yöntemler

İstikrarlılık, ‘Ölçüm aracı tekrar uygulandığında benzer sonuçlar üretiyor mu?’ sorusuna yanıt arar²⁴. Başka bir deyişle, istikrarlılık, tekrarlanan ölçümlerin benzerlik derecesi ile ilgilidir^{12,26}.

Eşdeğerlik, aynı ölçeğin iki paralel formunun aynı örnekleme art arda uygulanması ile ‘Aynı ölçüm aracının farklı sürümleri benzer sonuçlar verir mi?’ sorusuna yanıt arar^{24,26}.

İç tutarlılık, ‘Bir ölçüm aracındaki tüm maddeler aynı yapıyı ölçüyor mu?’ sorusu ile ilgilenir²⁴. Yani, ölçek içerisindeki maddeler kümesinin belirli bir kavram ya da özelliği ölçme derecesini sorgular^{23,29}.

Bir ölçek iç tutarlı ve/veya homojen olmalıdır yani, ölçeğin bütün alt birimleri aynı özelliği ölçmelidir^{12,26}. İç tutarlılık ve homojenlik kavramları arasındaki farkın bilinmesi önemlidir. İç tutarlılık maddelerin birbirleriyle bağıntılı/ilişkili olma derecesi ile ilgilenirken, homojenlik madde kümesinin tek boyutlu olma durumunu gösterir. Green ve ark. (1977), iç tutarlılığın homojenlik için gerekli fakat yeterli bir koşul olmadığını ifade etmişlerdir^{29,31,32}.

Güvenirlik, farklı şekillerde değerlendirilebilir. İstikrarlılık için test-tekrar test yöntemine, eşdeğerlik için paralel testler yöntemine, iç tutarlılık için ise yarıya bölme yöntemine başvurulur (Şekil 2.3.2.1). Bu yöntemlerden, paralel testler ve test tekrar test yöntemi testin aynı bireye birden fazla uygulanmasını gerektirirken, testin iki yarıya bölünmesi yönteminde tek bir uygulama yeterlidir. Ölçüm aracının bir kez uygulanması ile hesaplanabildiğinden, güvenilirlik çalışmalarında iç tutarlılık güvenilirliğine bağlı güvenilirlik katsayılarının hesaplanması en sık tercih edilen yöntemidir²³.

2.5.1. Test-Tekrar Test Yöntemi

Test-tekrar test güvenilirliği, zamansal istikrarlılık olarak tanımlanabilir²². Bir ölçme aracının güvenilirliğini tahmin etmenin bir yolu, aynı ölçüm aracının aynı katılımcı grubuna farklı zamanlarda uygulanmasıdır. Test-tekrar test güvenilirliği, aynı kişilerin aynı testin iki farklı uygulamasında elde edilen skorları arasındaki korelasyon değerinin

hesaplanmasına dayanır. Elde edilen bu değer, güvenilirlik katsayısını temsil etmektedir^{1,13,16,20,24}.

Test-tekrar test yöntemi, bir kişilik özelliği gibi zaman içerisinde nispeten kararlı olan bir şeyi ölçmek isteyen bir testin güvenilirliğini değerlendirirken uygundur. Zaman içinde ölçülen karakteristiğin durağan bir durum göstermeyeceği varsayılacak olursa, test-tekrar test yöntemini kullanarak testin güvenilirliğini değerlendirmek pek doğru olmaz. Zaman geçtikçe insanlar değişir. Örneğin, insanlar, yeni şeyler öğrenebilecekleri gibi bazı şeyleri de unutabilir. Genel olarak, aynı testin uygulamaları arasındaki zaman aralığı arttıkça, elde edilen skorlar arasındaki korelasyon azalır. Zaman geçişi bir hata varyansı kaynağı olabilir. Geçen süre ne kadar uzun olursa, güvenilirlik katsayısının düşük olma ihtimali o kadar yüksek olur^{16,23}. Fakat düşük test-tekrar test korelasyonu her zaman testin güvenilirliğinin de düşük olduğu anlamına gelmez. Bunun yerine altta yatan teorik kavramın değiştiğini gösterebilir. Örneğin, kişinin idam cezasına karşı tutumu cezayı almadan önce ve aldıktan sonra oldukça farklı olabilir. Ölçümler arası zaman aralığının uzaması, altta yatan kavramın değişmesi olasılığını artırır.

Test-tekrar test metodunun tipik bir problemi ilk uygulanan test ile tekrar testi arasındaki zaman aralığıdır. Örneğin, birinci ve ikinci test uygulamaları arasındaki zaman çok kısaysa, katılımcılar birinci test uygulamasından bir şeyler hatırlayabilirler ve böylelikle ikinci test uygulamasında vermiş oldukları cevaplar hafızalarından etkilenmiş olabilir. Eğer birinci ve ikinci test uygulamaları arasındaki zaman aralığı çok uzunsa, bilgi veya tutumların değişme olasılığı artar ve başlangıçtaki ölçümler ile daha sonra yapılan ölçümler arasında farklılık meydana gelir²³. Bu sebeple, katılımcıların orijinal yanıtlarını hatırlamayacakları kadar uzun olmalı, ancak materyalle ilgili bilginin değişmeyeceği kadar da kısa olmamalıdır. Genellikle, kabul gören zaman aralığı 2 hafta ile 1 ay arasında değişmektedir¹³.

Test-tekrar test metodunun bir diğer problemi, hafızaya bağlı aşırı tahmindir (overestimation). Genellikle, kişiler ikinci testin uygulanmasında birinci test uygulamasında verdikleri cevaplardan etkilenirler. Başka bir deyişle, ölçümler arası zaman kısaysa, denekler ilk başta verdikleri cevapları hatırlayabilirler ve bu durumda çalışmanın gerçekte olduğundan daha tutarlı gözükmesini sağlayabilir¹.

2.5.2.Paralel veya Alternatif Testler Yöntemi

Literatürde, sıklıkla birbirinin yerine kullanılabilir olmasına rağmen alternatif testler ve paralel testlerin birbirlerinden farklı olduğunu belirten kaynaklar da vardır. Bunlara göre: paralel testlerde, testin her bir formu için, gözlemlenen test puanlarının ortalamaları ve varyansları eşittir. Teorik olarak, paralel testlerde elde edilen puanlar, gerçek puan ile eşit derecede korelasyon gösterir. Alternatif testler ise, paralel olacak şekilde tasarlanmış bir testin biraz farklı versiyonlarıdır. Paralel test dizaynı gerekliliklerini karşılamazken, bir testin alternatif formları genellikle içerik ve zorluk seviyeleri derecelerine göre eşdeğer olacak şekilde tasarlanmıştır¹⁶.

Paralel veya alternatif testler yöntemi, ölçüm aracının farklı sürümleri ile ilgilidir^{13,24}. Bu yöntemde, iki uygulamada kullanılan ölçüm aracının paralel/alternatif sürümleri aynı kavramı ölçmeyi amaçlamalıdır. Ölçüm aracının farklı sürümleri aynı kişiye aynı veya farklı zamanlarda uygulanır. Güvenirlik tahminini sağlayan, ölçüm aracı ile ölçüm aracının paralel/alternatif sürümü arasındaki korelasyondur^{1,20}.

Ölçüm aracının farklı sürümlerinin uygulanma zamanları ile ilgili genelleştirilmiş bir bilgi yoktur. Önerilen, iki testin yaklaşık iki hafta arayla uygulanmasıdır¹. Ancak Nunnally ve Bernstein aynı gün testin her iki formunun da uygulanmasının uygun olabileceğini belirtmişlerdir. Brink and Wood (1998), ölçüm aracın her iki sürümünün aynı anda uygulanmasını önermiştir. Waltz ve ark. (2005), aynı zamanda uygulanan alternatif testlerin eşdeğerliği göstereceğini, istikrarın ise ölçüm araçları iki farklı zamanda uygulamak suretiyle elde edileceğini belirtmişlerdir¹³.

Alternatif testlerin veya paralel testlerin güvenilirliğine ilişkin tahminlerin elde edilmesi, iki yönden test-tekrar test yöntemiyle güvenilirliğin tahminine benzerdir:

- i. Aynı gruba iki testin iki defa uygulanması gerekmektedir. Fakat test-tekrar test yönteminde aynı testin aynı kişiye iki kez uygulanması söz konusu iken, alternatif/paralel testler yönteminde ilk uygulanan teste alternatif/paralel bir test ikinci uygulama için tercih edilir^{1,23}.
- ii. Test skorları, motivasyon, yorgunluk, uygulama şekli veya müdahale gibi faktörlerden etkilenebilir¹⁶.

Bir testin alternatif veya paralel bir biçiminin geliştirilmesinin, test kullanıcılarına çeşitli avantajları olmaktadır. Örneğin, hafızanın daha önce uygulanan test formunun içeriği üzerindeki etkisini alternatif/paralel test yöntemi ile en aza indirir¹⁶. Yani, alternatif/paralel testler yöntemi hafıza faktörünün güvenilirlik tahminini artırma olasılığını ortadan kaldırdığından test-tekrar-test yöntemine göre açıkça üstündür. Güvenirliğin değerlendirilmesinde kullanılan bu yöntemin temel kısıtlaması, paralel/alternatif testler oluşturma zorluğudur¹.

Alternatif formların güvenirliliği klinik araştırmalarda sıklıkla kullanılmaz ve madde havuzlarının genellikle daha büyük olduğu eğitim alanındaki çalışmalarda daha çok tercih edilir¹³.

2.5.3. Testin İki Yarıya Bölünmesi Yöntemi

Güvenirliliği değerlendirmede test-tekrar test ve paralel/alternatif test yöntemleri aynı kişilere aynı testin iki kez uygulanmasını gerektirmektedir. Yarıya bölme yöntemi ile elde edilen güvenilirlik tahmini ise, tek uygulama sonucunda testin eşdeğer yarımlarından elde edilen iki skor arasındaki korelasyonla elde edilir^{1,16,20}. Aslında, testi yarıya bölme konusundaki asıl amaç, iki paralel form oluşturulmasıdır. Böylelikle iç tutarlılık paralel formların güvenirliliğinin tahmini olarak da incelenebilir²⁴.

Testin iki yarıya bölünmesi yönteminin aşamaları aşağıdaki gibidir.

Adım 1: Test eşdeğer iki yarıya bölünür.

Elde edilen yarılar, eşit standart sapmaya sahip olmalıdır. Yani, iki yarım arasındaki korelasyonun, testteki maddelerin nasıl ayrıldığına dayalı olarak çok az değişmesi istenilendir. Eşit yarıları elde etmek için yapılan bölme işlemindeki herhangi başarısızlık, elde edilecek korelasyonun düşmesine sebep olacaktır. Bir testi bölmek için kabul edilebilir bir yol, maddeleri testin yarılarına rasgele atamaktır. Bir başka yol da, testin yarısına tek sayılı maddeleri ve diğer yarısına çift sayılı maddeleri atamaktır. Bu yöntem, aynı zamanda tek-çift güvenilirlik olarak da adlandırılan yarıya bölme güvenilirlik tahminini verir^{1,16,20}.

Adım 2: Eşdeğer iki yarı skorları arasındaki Pearson korelasyon değeri hesaplanır.

Adım 2 temel bir sıkıntıyı içerir. Hesaplanan Pearson korelasyon değeri total test için değil de her bir yarı için güvenirlik değerine denktir. Yani, ölçeği yarıya bölmek, ölçek uzunluğunu %50 kısaltacağından, güvenirlik değerinin olduğundan düşük tahmin edilmesine neden olur. Bu sorun, Speraman-Brown formülü kullanılarak kolayca çözülebilir^{1,16,33}.

Adım 3: Spearman-Brown formülü kullanılarak güvenirlik değeri düzeltilir.

Spearman-Brown formülünün genel hali,

$$r_{SB} = \frac{nr_{xy}}{1+(n-1)r_{xy}} \quad (6)$$

şeklindedir. Eşitlik (6)' da,

r_{SB} : Spearman-Brown güvenirliği

r_{xy} : orjinal test için Pearson korelasyon değeri

n : orijinal testin yarıya bölünme sayısı olarak tanımlanır.

Yarıya bölme yönteminde, yarımlardan biri için güvenirlik katsayısı hesaplandıktan sonra, Spearman-Brown formülü kullanılarak tüm test için güvenirlik katsayısı hesaplanabilir. Tüm test, yarım testten 2 kat daha uzun olduğundan Spearman-Brown formülü n yerine 2 yazılarak yarıya bölme güvenirliğine uyarlanır.

$$r_{SB} = \frac{2r_{hh}}{1+r_{hh}}$$

r_{hh} : iki yarı arasındaki Pearson korelasyon değeri

Yarıya bölme metodunda güvenirlik 0 ile 1 arasında değer alır. Tüm testin ele alındığı durumdaki güvenirlik tahminleri, bir testin yarısına dayalı olan tahminlerden daha yüksektir^{1,4,16,21}.

Yarıya bölme metodunun, test-tekrar test ve paralel/alternatif formlardan daha kullanışlı olmasını sağlayan çeşitli yönleri vardır. Bunlardan ilki, hafızanın etkisinin bu

metotta ortadan kalkmasıdır. Bir diğeri ise, yarıya bölme metodunun uygulanması genellikle daha ucuzdur ve testin iki yarısı aynı anda uygulandığından zamana bağlı yöntemlerle kıyaslandığında sonuçların daha kolay elde edildiği görülmektedir. Nunnally(1978) , bir testin alternatif formu mevcut olmadığı durumda ve kısa zaman periyodu içerisinde davranışsal değişiklikleri ölçmek için yarıya bölme metodunun kullanılmasını önermiştir²³.

Yarıya bölme yöntemi, iç tutarlılığı ölçmek için kullanılan ilk metottur. İki testle güvenilirliği değerlendirmek veya iki kez test uygulamak bazı durumlarda zor olduğundan, yarıya bölme metodu güvenilirlik için kullanışlı olur¹⁶. Ancak, ölçekleri yarıya bölmek için birçok yol olması ve her birinin farklı güvenilirlik tahminleri vermesi önemli bir sorundur³³. G. F. Kuder ve M. W. Richardson (1937) tarafından bu sorunun üstesinden gelmek için ikili (dichotomous) maddeler söz konusu olduğunda Eşitlik (7) önerilmiştir^{1,4,33}.

$$KR_{20} = \frac{k}{k-1} \left[1 - \frac{\sum p_k q_k}{\sigma_{toplaml}^2} \right] \quad (7)$$

Eşitlik (7)'de,

k : madde sayısı

p_k : k. maddeyi pozitif cevaplayan kişilerin oranı

q_k : k. maddeyi negatif cevaplayan kişilerin oranı

$\sigma_{toplaml}^2$: toplam skor varyansı temsil etmektedir.

KR_{20} 'nin temel varsayımı, maddeler arası korelasyon matrisinin rankının 1 olmasıdır. Bu, tüm maddelerin aynı faktörü ölçtüğü duruma karşılık gelir^{20,21}.

KR_{20} , mümkün olan bütün yarıya bölme güvenilirlik değerlerinin ortalaması gibi düşünülebilir ve eğer test maddelerinin homojenliği yüksekse, KR_{20} ve yarıya bölme yöntemi güvenilirlik tahminleri benzerdir. Eğer test maddelerinin heterojenliği yüksekse, KR_{20} güvenilirliği yarıya bölme metodundan daha düşük tahmin eder^{16,20,21}.

Testin iki yarıya bölünmesi yönteminde, tüm teste ilişkin güvenirliliğin belirlenmesinde kullanılan bir diğer yaklaşım Guttman yöntemidir. Guttman çeşitli durumlarda güvenirliliği belirlemek için 6 farklı alt sınır elde etmiştir. Bütün alt sınırların ortak özelliği, tahminin tek denemeden elde edilmesidir. Formülasyonlar farklı olmasına rağmen Guttman katsayılarından L_3 ile elde edilen tahmin KR_{20} katsayısına denktir. Madde kovaryanslarını kullanan L_2 orijinal formüldür ve diğer 5 katsayıya göre hesaplanması daha zordur. L_4 ise yarıya bölme katsayısına denktir. Guttman'ın, yarıya bölme katsayısı,

$$L_4 = 2 \left(1 - \frac{s_1^2 + s_2^2}{s_t^2} \right) \quad (8)$$

şeklindedir. Eşitlik (8)'de,

s_1^2 : testin ilk yarısına ilişkin varyans

s_2^2 : testin ikinci yarısına ilişkin varyans

s_t^2 : testin iki yarısının toplanması ile elde edilen değişkenin varyansdır.

Eşitlik (8)'de verilen Gutman katsayısı, iki test yarısının eşit varyansa sahip olması gibi bir varsayım bulundurmadığından Spearman-Brown formülünden daha üstün olduğuna dikkat çekilir^{5,20,21}.

Testin iki yarıya bölünmesi yönteminde kullanılan güvenirlilik katsayısı hesaplama yöntemlerinden biri olan KR_{20} 'nin sadece ikili maddelerle sınırlı kalma problemi, formülünün genelleştirilmesi ihtiyacını doğurmuştur^{16,33}.

2.6. Bir Güvenirlilik Ölçütü olarak Cronbach'ın Alfası

Güvenirliliğin önemli bileşenlerinden olan iç tutarlılığı değerlendirmek için kullanılan yarıya bölme metodu birçok konuda eleştirilir. Bu eleştirilerin başında, bu metotta, ölçek için tek bir katsayı vermek yerine, ölçeğin iki bölüme ayrılmasıyla hangi maddelerin gruplandırıldığına bağlı olarak farklı katsayılar elde edilmesi gelmektedir³⁴. Örneğin, 12 maddeli bir ölçek, 462 farklı yarıya bölünebilir ve her bölme işlemi sonucu farklı güvenirlilik tahmini yapılır¹¹. Bu probleme ilk çözüm Kuder-Richardson(1937)

tarafından önerilmiştir. Ancak, önerilen KR_{20} katsayısı sadece iki yanıt seçenekli maddelerle sınırlı kalmıştır. Sadece ikili maddelerle sınırlı kalma problemi ise KR-20 formülünün genelleştirilmesi ile Cronbach(1951) tarafından çözülmüştür^{4,33}.

Güvenirlik değerlendirmelerinde en çok iç tutarlılığa başvurulur. İç tutarlılığı değerlendirmek için de kullanılan en yaygın metot Cronbach'ın alfa katsayısıdır^{1,10,11,29,32}. En yaygın metot olmasının başlıca sebebi, ölçeğin iki kez uygulanmasını ya da iki veya daha çok değerlendirici gerektirmemesi ve böylece test-tekrar test yöntemine veya değerlendiriciler arası güvenilirliğe göre daha kolay hesaplanabilir olmasıdır^{4,33}.

Literatürde alfa katsayısı ile ilgili farklı tanımlamalar vardır:

- Alfa katsayısı, bir testin tüm mümkün yarıları arasındaki korelasyonun ortalamasının Spearman-Brown formülü ile düzeltilmiş hali olarak düşünülebilir^{16,33}. Yani, mümkün bütün yarıya-bölme güvenilirliklerinin ortalaması alfa katsayısıdır^{4,14,33,34}.
- Alfa katsayısı, KR_{20} katsayısının genel versiyonu olarak tanımlanabilir^{4,24,32}.
- Alfa katsayısı, gerçek bir test ile aynı uzunluğa sahip varsayımsal bir alternatif test arasındaki beklenen korelasyonun bir tahmini olarak düşünülebilir¹.
- Alfa katsayısı, tau-eşdeğer durumundaki güvenirliliktir^{11,29,32}.
- Literatürde, iç tutarlılık genellikle güvenirliliğin alt sınırı olarak tanımlanmasına rağmen alfa katsayısının güvenirliliğin alt sınırı veya üst sınırı olarak tanımlanması konusunda bazı tartışmalar bulunmaktadır. Alfa, paralel veya tau-eşdeğer ölçekler için güvenirliliğin üst sınırı, konjenerik ölçekler içinse güvenirliliğin alt sınırı olarak bilinir⁴.

Literatürde alfa katsayısının hesaplanmasına yönelik farklı yaklaşımlar da mevcuttur. Alfa katsayısı genel olarak,

$$\alpha = \frac{k}{k-1} \left[1 - \frac{\sum \sigma_i^2}{\sigma_{toplaml}^2} \right] \quad (9)$$

eşitliği ile verilir. Eşitlik (9)'da,

k : ölçekte bulunan madde sayısı

σ_i^2 : ölçekte bulunan i . maddenin varyansı

$\sigma_{toplamlam}^2$: toplam varyans

şeklinde tanımlanır^{4,34}. Toplam varyans ise,

$$\sigma_{toplamlam}^2 = \sum_{i=1}^k \sigma_i^2 + \sum_{i \neq j} \sum \sigma_{ij} \quad (10)$$

eşitliği ile elde edilir. Eşitlik (10)'da, σ_{ij} : i . ve j . maddeler arasındaki kovaryanstır¹⁰.

Eşitlik (9)'da verilen alfa katsayısı, matris gösterimi ile ifade etmek istenilirse, $\mathbf{X} = [X_1 \dots X_k]$, k maddeli bir ölçeğin yanıt değişkeni olsun. Alfa katsayısını matris gösterimi,

$$\alpha = \frac{k}{k-1} \left[1 - \frac{tr(\mathbf{\Sigma})}{\mathbf{1}'\mathbf{\Sigma}\mathbf{1}} \right] \quad (11)$$

şeklindedir. Eşitlik (11)'de, $\mathbf{\Sigma}$ popülasyon varyans-kovaryans matrisidir ve $\mathbf{1}$, $k \times 1$ boyutlu 1'lerden oluşan bir vektördür³⁵.

Eğer varyans-kovaryans matrisi yerine korelasyon matrisi ele alınıyorsa, Cronbach'ın alfası aşağıdaki biçime indirgenir:

$$\alpha = k\bar{\rho}/[1 + \bar{\rho}(k - 1)] \quad (12)$$

Eşitlik (12)'de,

$\bar{\rho}$: maddeler arası ortalama korelasyon

k : ölçekte bulunan madde sayısı şeklinde tanımlanır.

Maddeler arası ortalama korelasyon arttıkça/azalmadıkça ve madde sayısı arttıkça alfa değerinin yükseldiği Eşitlik (12)'den görülür^{1,5}.

Alfa katsayısının hesaplanmasına ilişkin yaklaşımlardan biri de tekrarlı ölçümlerde tek yönlü varyans analizine dayanır. Bu yaklaşıma göre alfa katsayısı,

$$\alpha = \frac{DAKO-HKO}{DAKO} \quad (13)$$

ile verilir. Eşitlik (13)'te,

$DAKO$: Denekler/gözlemler arası kareler ortalaması

HKO: Hata kareler ortalaması

ile ifade edilir⁵.

2.6.1. Alfa Katsayısının Yorumlanması

Cronbach alfa katsayısı 0 ile 1 arasında değer alır^{1,16}. Alfa katsayısının büyüklüğü, hem ölçekte bulunan madde sayısından hem de ortalama maddeler arası korelasyondan etkilenir¹¹. Düşük alfa değerleri, ölçekte maddeler arasında korelasyon eksikliğini gösterirken, yüksek alfa değerleri ise, ölçeğin maddelerindeki yüksek korelasyonları gösterir. Ayrıca, alfa katsayısı ölçekteki madde sayısına bağlı olduğundan çok sayıda madde içeren ölçekler için çok yüksek alfa değeri bulunur. Yani, yüksek alfa değerleri, ilişkili oldukları sürece maddeler eklenerek elde edilebilir^{1,12}. Ölçekteki madde sayısı ve ortalama maddeler arası korelasyonlarına ek olarak, skor varyanslarındaki farklılık da bir ölçüm aracından elde edilen skorların güvenilirliğini, yani alfa katsayısını etkiler^{11,32}.

Alfanın yüksek değerleri her zaman istenilendir. Fakat bir ölçeğin alfa değeri 0.90'dan yüksek ise gereksiz maddeler barındırma olasılığının yüksek olduğunu söylenir. Yani, aynı soruyu biraz farklı şekillerde soran bir takım madde varlığı söz konusudur. Bu durum, bazı maddelerin ölçekten çıkarılması gerektiğini ve ek olarak ölçeğin bir bütün olarak kapsamının çok dar olabileceğini (ki bu çok yüksek geçerliğe sebep olur) gösterir^{4,16,29}.

Kabul edilir alfa seviyesi için belirlenmiş bir düzey veya sınıflandırma yoktur, çalışmadan çalışmaya değiştiği söylenir³¹. Genel olarak güvenilirlik katsayılarının 0.80'den düşük olmaması istenmektedir¹. Nunnally ve Bernstein, iç tutarlılık için 0.70-0.90 değer aralığını iç tutarlılığın iyi bir ölçüsü olarak önermiştir. Nunnally (1978), 0.70 değerinin araştırmanın ilk aşamalarında elde edilmesi gerektiğini, ancak temel araştırmalarda en az 0.80 olması ve klinik bir ölçüm aracı için ise 0.90 olması gerektiğini belirtmektedir³⁶. Ponterotto ve Ruckdeschel tarafından alfanın, ölçeğin uzunluğundan ve örneklem büyüklüğünden etkilendiğini göz önüne alan biraz farklı bir yaklaşım öne sürülür ve bu faktörlere bağlı olarak bir dizi değer önerilir. Örneğin, yediden az sayıda maddeden oluşan ve 100 gözlem alınarak değerlendirilen bir ölçek için 0.70 alfa değeri 'iyi' sınıfına girmektedir. Bununla birlikte, ölçekteki madde sayısı 11'den fazlaysa ve

örnek büyüklüğü 300'ün üzerindeyse, α 'nın 'iyi' sınıfına girmesi için 0.90 olması gereklidir⁴.

Cronbach'ın alfası, tek boyutluluk veya homojenlik için bir ölçüt değildir^{11,31}. Temel olarak güvenilirlik kavramı, test maddelerinin örnekleminin tek boyutlu olduğunu varsayar ve bu varsayımın ihlal edilmesi güvenirliliğin önemli derecede düşük tahminine neden olur. Alfa katsayısı, madde etkileşiminin bir fonksiyonu olarak artarken, çok boyutluluğun bir fonksiyonu olarak azalan bir yapı sergiler³². Çok boyutluluk durumu, alfa katsayısı ile güvenilirlik tahmininin olduğundan düşük elde edilmesine yol açar³¹. Alfa değeri belirli bir madde ölçek dışı bırakıldığında belirgin bir şekilde artarsa, bu durum dışlamanın ölçeğin homojenliğini artıracağını gösterir fakat alfa katsayısının homojenlik ya da tek boyutluluk için bir indeks olduğunu göstermez⁴.

Çok boyutluluğa ek olarak, maddeler arası korelasyon matrisinin rankı 1 olmadığı durumlarda da alfa katsayısı güvenirliliği olduğundan az tahmin edebilir. Maddeler arası korelasyon matrisinin rankının 1 olması durumunda ise alfa katsayısı güvenirliliğin kesin bir tahminidir. Aksi takdirde, tutucu bir tahmin elde edilir^{20,31}.

Ölçek birden fazla kavramı veya yapıyı barındırıyorsa, alfa katsayısının değerini arttırmak için daha fazla soru kaçınılmaz olacağından, testin tamamı için Cronbach alfa değerini vermek mantıklı olmayabilir. Bu nedenle, homojen alt ölçeklerin sayısını belirledikten sonra, alfa katsayısı her alt ölçek için ayrı ayrı hesaplanmalıdır^{12,29}.

Düşük güvenilirlik, anlamlı ilişkileri tanımlamak için istatistiksel testlerin gücünü azaltır ve bu nedenle tip 2 hata olasılığını artırır. Yani düşük güvenilirlik, reddedilmesi gerektiğinde, sıfır hipotezini reddetmeme olasılığını artmasına neden olur. Bu da, araştırmacılar tarafından istenilen bir durum değildir¹¹.

2.6.2. Alfa Katsayısı için Güven Aralığı

Ölçeklerde alfa katsayısı, tau-eşdeğer olan k madde toplamının güvenirliliğini tanımlamak için kullanılmasına rağmen, α için klasik güven aralığı, paralel ölçümler ve çok değişkenli normal dağılım gibi güçlü varsayımlar barındırmaktadır⁷. α , daha az

sınırlayıcı “tau-eşdeğer” ölçüm varsayımı altında uygun güvenilirlik ölçütü olduğundan dolayı, kullanılan klasik güven aralığı oldukça sınırlayıcıdır⁸.

Eşitlik (11), çok değişkenli normallik varsayımına dayanan alfa katsayısının en çok olabilirlik tahminidir. Van Zyl, Neudecker ve Nell(2000), ölçümlerin genel kovaryans yapısının sınırlayıcı olmayan varsayımı altında, alfa katsayısının tahmin edicisi için yaklaşık varyans elde etmişlerdir. Bu varsayımla, $n \rightarrow \infty$ iken $\sqrt{n}(\hat{\alpha} - \alpha)$, 0 ortalamalı Q varyanslı normal dağılıma sahiptir($\hat{\alpha}$: α ’nın en çok olabilirlik tahmini)³⁷.

$$Q = \left[\frac{2k^2}{(k-1)^2(j'Vj)^3} \right] [(j'Vj)(trV^2 + tr^2V) - 2(trV)(j'V^2j)] \quad (14)$$

n :örnek büyüklüğü

V : madde kovaryans matrisi

Alfa katsayısı için $(1 - \alpha)\%$ güven aralığı aşağıdaki gibi elde edilmiştir.

$$1 - \exp[\ln(1 - \hat{\alpha}) - \ln[n/(n - 1)] \pm z_{\alpha/2} \sqrt{var(\hat{\alpha})/(1 - \hat{\alpha})^2}] \quad (15)$$

Eşitlik (15)’te,

$$\hat{\alpha} = [k/(k - 1)][1 - tr(V)/j'Vj]$$

$\ln[n/(n - 1)]$: yanlılık düzeltmesi

$j : k \times 1$ vektördür.

Güven aralığında kullanılan varyans formülü, Van Zyl et.al(2000) tarafından elde edilen varyans formülünde(Eşitlik (14)) n yerine $n - 3$ alınmasıyla ufak bir değişiklikle Eşitlik(16)’daki gibi ifade edilir.

$$var(\hat{\alpha}) = \left[\frac{2k^2}{(k-1)^2(j'Vj)^3} \right] [(j'Vj)(trV^2 + tr^2V) - 2(trV)(j'V^2j)]/(n - 3) \quad (16)$$

Alfa katsayısı için klasik olan güven aralığının aksine, eşitlik (15) ; hem konjenerik hem de tau-eşdeğer ölçümler için uygundur. Ancak α ‘nın konjenerik ölçümlerin güvenilirliğini olduğundan az tahmin ettiği bilinmelidir. Klasik güven aralığı için de (15)

numaralı denklemde verilen güven aralığı için de ölçümlerin çok değişkenli normal dağılıma yakınsadığı varsayılmaktadır⁸.

Paralel ölçümler için ise, Bonett tarafından önerilen aşağıdaki yaklaşık güven aralığı kullanılabilir³⁸.

$$1 - \exp \left[\ln(1 - \hat{\alpha}) - \ln\left\{\frac{n}{n-1}\right\} \pm z_{\alpha/2} \sqrt{2k/[(k-1)(n-2)]} \right] \quad (17)$$

Eşitlik (17)'de $2k(1 - \hat{\alpha})^2/[(k-1)(n-2)]$, $var(\hat{\alpha})$ 'nın tahmini olarak kullanılır.

Cronbach alfanın hesaplandığı çalışmalarda, rutin olarak eşitlik (15)'in kullanımı önerilmektedir fakat eşitlik (17)'nin hesaplamasal kolaylığından yararlanmak istenirse örnek büyüklüğü formülü için paralel ölçümlerde $var(\hat{\alpha}) = 2k(1 - \hat{\alpha})^2/[(k-1)(n-2)]$ kullanılabilir⁸.

Cronbach Alfa Katsayısı için Hipotez Testi Yaklaşımı: h , araştırmacı tarafından belirlenen bir değer olmak üzere, $H_0: \alpha = h$ hipotezinin testi, aşağıdaki 3 karar kuralı kullanılarak Eşitlik (15)'ten elde edilmektedir.

- Eğer güven aralığının alt limiti h değerinden büyükse, H_0 reddedilir ve $\alpha > h$ kabul edilir,
- Eğer güven aralığının üst limiti h değerinden küçükse, H_0 reddedilir ve $\alpha < h$ kabul edilir,
- Bu iki durum dışındaki sonuçlar ise ikna edici değildir.

Cronbach alfa değeri güven aralığı ile birlikte verilmelidir. Çünkü hipotezi test ederken literatürde belirli bir h sabit değeri yoktur fakat genel olarak 0.70 tercih edilen değerdir¹⁰.

2.6.3. Güvenirlik Çalışmalarında Alfa Katsayısına Bağlı Örnek Büyüklüğü

Sağlık alanında yapılan araştırmalara, deneklerden kaynaklanan örnekleme hatasını en aza indirmek için yeterli sayıda kişi dahil edilmelidir. Hatayı en aza indirmek sadece klasik test teorisi ile değil aynı zamanda yeterli örnek büyüklüğü ile de sağlanmalıdır.

Çünkü klasik test teorisi araştırmalarda küçük örnek büyüklüğü kullanılmasından kaynaklanan hatayı gidermede yeterli değildir⁹.

Örnek büyüklüğü belirlenmesi belki de güvenilirlik çalışmalarının dizaynında en önemli konulardan biridir. Eğer örnek büyüklüğü çok küçük olursa, testin gücü düşük ve güven aralığı çok geniş olabilir. Örnek büyüklüğünün çok fazla geniş olması ise kaynak israfına sebep olabilir. Verilerin toplanması konusunda finansman talep eden araştırmacılar, genellikle önerilen örnek büyüklüğünün ne çok küçük ne de çok büyük olmaması gerektiğine dikkat etmelidir. Gerçektir ki, örneklem genişliği arttıkça, alfa katsayısı tahminleri yaklaşık olarak hatasız elde edilebilir. Literatürde, güvenilirlik çalışmaları için gerekli örnek büyüklüğü çok değişkenlik göstermektedir^{7,35}.

İç tutarlılık çalışmalarında, yüksek doğrulukla güvenilirlik elde etmek için örnek büyüklüğünün küçük olmaması gereklidir. Örnek büyüklüğünün küçük olması, güvenilirlik katsayıları ile beraber güven aralıklarının da yanlış tahmin edilmesine sebep olur⁶. Güven aralığı genişliği ne kadar azsa, güvenilirlik tahmini o kadar doğrudur, kararlıdır^{9,39}.

Cronbach alfa katsayısına bağlı örnek büyüklüğü formülleri, eşit varyanslara veya eşit kovaryanslara gereksinim duyulmayan bir güven aralığı yaklaşımı ile istenilen doğruluğa ve hipotez testlerinde istenilen gücü elde etmeye bağlı olarak 2 çeşittir. Bu formüller, örnek büyüklüğünü yaklaşık olarak belirlemek adına kullanılır⁸.

2.6.3.1. İstenilen Doğruluk için Örnek Büyüklüğü Hesaplama Yaklaşımı

Amacı alfa katsayısı için güven aralığını belirlemek olan çalışmalarda, araştırmacılar, istenilen güven aralığı genişliği ve istenilen güven düzeyini sağlayan örnek büyüklüğü kullanırlar. $\tilde{\alpha}$, Cronbach alfa güvenilirlik için planlanan değer olsun. Bu planlanan değer, uzman görüşlerinden, pilot çalışmadan veya literatür incelenmesinden elde edilebilir. İstenilen genişlik (w) ile α için $(1 - \alpha)\%$ güven aralığı elde etmede gerekli örnek büyüklüğü için, ilk olarak aşağıda verilen ön örnek büyüklüğü yaklaşımı elde edilir.

$$n_0 = [8k/(k - 1)](1 - \tilde{\alpha})^2(z_{\alpha/2}/w)^2 + 2 \quad (18)$$

Burada, $z_{\alpha/2}$ iki yönlü kritik z değeridir. Ön örnek büyüklüğü formülü eşitlik (18) kullanılarak ve $\hat{\alpha}$ yerine $\tilde{\alpha}$ konularak Eşitlik (17)'deki güven aralığı hesaplanır. Bu güven aralığının genişliği (w_0) hesaplanır ve Bonett ve Wright 'in örnek büyüklüğü yaklaşımı Eşitlik (19) ile verilir^{7,8}.

$$n = (n_0 - 2) (w_0/w)^2 + 2 \quad (19)$$

Eşitlik (19)'un güven aralığı genişliğini barındırması, paralel ve tau-eşdeğer/konjenerik ölçümlere dayalı güven aralığı yaklaşımlarının farklı olmasından dolayı farklı gereksinimler doğurur:

- Paralel ölçümler için örnek büyüklüğü yaklaşımında, araştırmacılara sadece Cronbach alfa güvenirlik katsayısı için planlanan değerin belirlenmesinde gereksinim duyulur.
- “tau-eşdeğer” ve “konjenerik” ölçümler için örnek büyüklüğü gereksinimleri kolaylıkla elde edilir fakat bu örnek büyüklüğü yaklaşımı, k tane varyans ve $k(k - 1)/2$ tane kovaryansın planlanan değerlerinin belirlenmesini gerektirir. Bu planlanan değerler bilinmeyen popülasyon varyansı ve kovaryansı ile denk düşmezse, örnek büyüklüğü yaklaşımının doğruluğu düşebilir.

2.6.3.2. İstenilen Güç için Örnek Büyüklüğü Hesaplama Yaklaşımı

Amacı alfa katsayısının değeri ile ilgili olan hipotezi test etmek olan çalışmalarda, etki büyüklüğü ve bazı belirlenmiş α değerleri için istenilen güç düzeyini sağlayan örnek büyüklüğü kullanılır. Hipotez testi uygulamalarında, daha güçlü testler genellikle büyük örnek büyüklükleri ile elde edilmektedir.

İstenilen gücü elde ederek, $H_0: \alpha = h$ hipotezini test etmek için gerekli örnek büyüklüğü yaklaşık olarak aşağıdaki gibidir⁸.

$$n = [2k/(k - 1)](z_{\alpha/2} + z_{\beta})^2/(\tilde{\alpha}^* - h^*)^2 + 2 \quad (20)$$

Eşitlik (20)'de,

$$\tilde{\alpha}^* = \ln(1 - \tilde{\alpha})$$

$$h^* = \ln(1 - h)$$

z_β : tek yönlü kritik z değeri ve

$\beta = 1 - \text{güç}$ şeklinde ifade edilir.



3. GEREÇ ve YÖNTEM

Simülasyon Çalışmaları

Simülasyon çalışmalarındaki amaç, farklı senaryolarla güvenilirlik çalışmalarında istenen doğruluk ve istenen güç için elde edilen örnek büyüklüklerini değerlendirmektir.

Simülasyon Çalışması 1: İstenilen doğruluk için örnek büyüklüğü elde etme üzerine kurulu ilk simülasyon çalışmasında, paralel ölçüm varsayımı göz önüne alınarak, planlanan farklı Cronbach alfa değer aralıkları, ölçekte bulunabilecek farklı madde sayıları ve istenen farklı güven aralığı genişlikleri altında örnek büyüklükleri incelenmiştir. Burada, planlanan cronbach alfa değer aralıkları,

- 0,60-0,69 ve 0,70-0,79 “orta güvenilirlik”
- 0,80-0,89 ve 0,90-0,99 “yüksek güvenilirlik”

baz alınarak rastgele üretilmiştir. Düşük güvenilirlik aralığı çalışmalarda istenen seviyeyi yansıtmadığından simülasyon çalışması kapsamında ele alınmamıştır. Ölçekte bulunan madde sayıları 5, 10, 20, 36 (uygulamalarda sıklıkla kullanılan SF-36 ölçek verisi değerlendirmeye alındığı için), 50, 75 ve 100, istenilen güven aralığı genişliği 0.05, 0.1 ve 0.15 olarak alınmış ve 1000 tekrarlı simülasyonla örnek büyüklüğü yaklaşımı değerlendirilmiştir.

Simülasyon Çalışması 2: İstenilen doğruluk için örnek büyüklüğü elde etme üzerine kurulu ikinci simülasyon çalışmasında Simülasyon Çalışması 1’e ek olarak, planlanan cronbach alfa değer aralıklarının, “orta güvenilirlik” ve “yüksek güvenilirlik” sınıfında olması kısıtı altında 1000 varyans-kovaryans matrisi elde edilmiştir. Başlangıçta, boyutu ölçekte bulunan madde sayısına bağlı olan varyans-kovaryans matrislerinin elamanları 0.1-0.99 aralığında düzgün dağılımdan üretilmiştir. Daha sonra, varyans-kovaryans matrislerinin,

- Köşegen dışı elemanları (kovaryanslar), planlanan güvenilirlik kısıtını sağlaması göz önüne alınarak üretilen değerlerden rastgele seçilip karesi alınarak,

- Köşegen elemanları (madde varyansları) ise, $0.6^2-0.9^2$ arasında düzgün dağılımdan üretilen hata değerlerinin rastgele seçilen değere eklenmesi ile elde edilmiştir.

Simülasyon Çalışması 3: İstenilen güç için örnek büyüklüğü elde etme üzerine kurulu simülasyon çalışmasında ise, planlanan farklı Cronbach alfa değer aralıkları ve bu aralıklara bağlı olarak $H_0: \alpha = h$ hipotezini test etmek adına farklı h değerleri dikkate alınarak incelenmiştir. Sıfır hipotezinde yer alan h değerleri planlanan Cronbach alfa değer aralıklarının alt limitlerinden düşük olacak şekilde seçilmiş ve aşağıda verilmiştir.

- 0.60-0.69 iken; $h=0.50$
- 0.70-0.79 iken; $h=0.60$
- 0.80-0.89 iken; $h=0.70$
- 0.90-0.99 iken; $h=0.80$

Ölçekte bulunan madde sayıları Simülasyon Çalışması 1 ile aynı iken güç değerleri 0.80, 0.85 ve 0.90 olarak alınmıştır.

Gerçek Veri Çalışmaları

Oswestry Bel Ağrısı Ölçeği (The Oswestry Disability Index-ODI): Çalışmada kullanılan gerçek veri setlerinden ilki, Kasım 2015 - Ağustos 2016 tarihleri arasında ODI'nin Çukurova Üniversitesi Tıp Fakültesi Balcalı Hastanesinde görev yapan ve bel ağrısı bulunan 212 hemşireye uygulanması ile elde edilmiştir⁴⁰. ODI, bel ağrılarının kaynaklanan fonksiyonel yetersizliği ve bel ağrısının günlük hayat üzerindeki etkilerini değerlendirmek amacıyla geliştirilmiştir. Bu ölçek, ağrının şiddeti, kişisel bakım, kaldırma, yürüme, oturma, ayakta durma, uyuma, ağrının değişme derecesi, cinsel yaşam ve seyahat olmak üzere 10 maddeden oluşan tek boyutlu bir ölçektir. Her madde 0 - 5 arasında puan alan 6'lı likert yanıt seçeneğinden oluşmaktadır. Ölçek puanı 0 ile 50 arasında değişmektedir⁴¹.

İstenilen doğruluk ve istenilen güç için yapılan çalışmalarda, Cronbach alfa değeri, Cronbach alfaya dayalı güven aralığı ve güç hesaplamaları gerçek veriden rastgele 1000 tekrarlı 20, 40, 80, 120, 160 ve 200 örnek çekilerek yapılmıştır.

Yaşam Kalitesi Ölçeği (SF-36): Çalışmada kullanılan gerçek veri setlerinden ikincisi, Çukurova Üniversitesi Tıp Fakültesi Balcalı Hastanesi'nin Fizik Tedavi ve Rehabilitasyon Anabilim Dalı'nda, 120 skleroderma hastasına SF-36 ölçeğinin uygulanması ile elde edilmiştir. SF-36, sağlık alanında, bireylerin yaşam kalitesini değerlendirmede kullanılan ölçeklerden biridir ve 36 maddeden oluşmaktadır. Fiziksel fonksiyon (FF), fiziksel rol güçlüğü (FRG), emosyonel rol güçlüğü (ERG), vitalite (V), vücut ağrısı (VA), genel sağlık algısı (GSA), sosyal işlevsellik (Sİ) ve ruhsal sağlık (RS) olmak üzere 8 alt boyutu (ölçeği) barındırmaktadır. Bu alt ölçeklerden FF 10 maddeden, RS ve GSA 5 maddeden, FRG ve V 4 maddeden, ERG 3 maddeden ve son olarak VA ve Sİ 2 maddeden oluşmaktadır. Her alt ölçek için ayrı ayrı puanlar hesaplanabilir ve bu puanlar 0-100 arasında değer alır. Yüksek puanlar genel sağlık durumuyla ilgili yaşam kalitesinin daha yüksek olduğunu göstermektedir. Ölçek için genel bir toplam puan hesaplanması söz konusu değildir⁴².

İstenilen doğruluk ve istenilen güç için yapılan çalışmalarda, Cronbach alfa değeri, Cronbach alfaya dayalı güven aralığı ve güç hesaplamaları gerçek veriden rastgele 1000 tekrarlı 10, 20, 40, 80 ve 120 örnek çekilerek yapılmıştır.

Verilerin analizinde SPSS versiyon 23 ve Rstudio versiyon 1.0.143 paket programlarından yararlanılmıştır^{43,44}.

4. BULGULAR

4.1.Simülasyon Çalışmasına ait Sonuçlar

Simülasyon çalışmasına ait sonuçlar,

- İstenilen doğruluğa dayalı örnek büyüklüğü için simülasyon çalışmaları
- İstenilen güce dayalı örnek büyüklüğü için simülasyon çalışmaları

olmak üzere iki başlık altında toplanmıştır.

4.1.1. İstenilen Doğruluğa Dayalı Örnek Büyüklüğü için Simülasyon Sonuçları

4.1.1.1. Paralel ölçüm varsayımı ile elde edilen simülasyon sonuçları

Paralel ölçüm varsayımı altında, istenilen doğruluğa dayalı örnek büyüklüklerine ait simülasyon çalışması sonuçları sırasıyla istenilen güven aralığı genişliklerinin 0.05, 0.1 ve 0.15 olduğu durumlarda incelenmiştir.

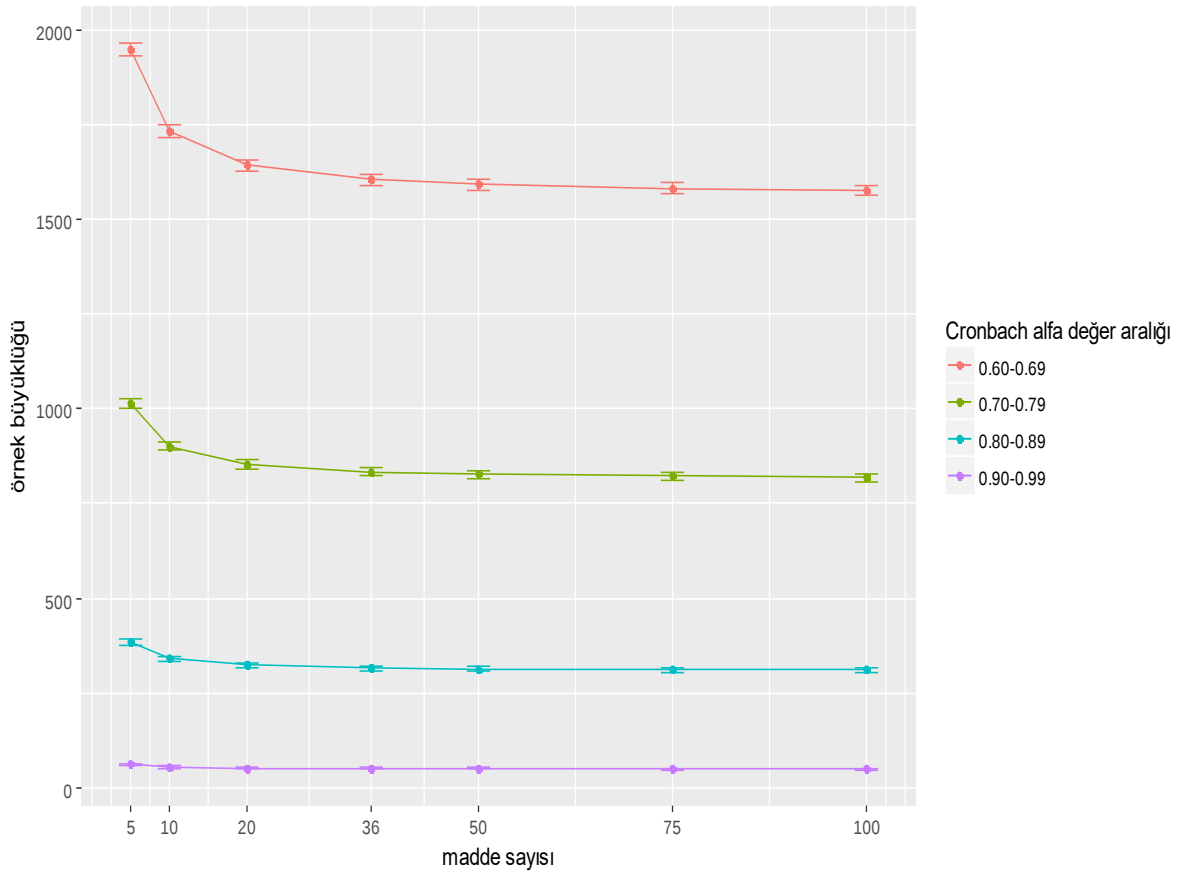
İstenilen güven aralığı genişliğinin 0.05 olduğu senaryo ile ilgili bulgular Çizelge 4.1.1’de, grafik gösterimi ise Şekil 4.1.1’de verilmiştir. Ölçekte bulunan madde sayısı 5 olduğu durumda planlanan Cronbach alfa aralığı dikkate alındığında, güvenilirlik değerleri yükseldikçe örnek büyüklüğünün önemli düzeyde düştüğü görülmüştür. Bu durumu genelleştirecek olursak, farklı madde sayılarının alındığı durumların hepsinde planlanan yüksek güvenilirlik düzeyine ait örnek büyüklüğü ortalamasının daha düşük olduğu görülmüştür. Planlanan Cronbach alfa değer aralığının 0.60-0.69’dan 0.90-0.99’a çıkması her farklı madde senaryosunda örnek büyüklüğünde %3’lük bir azalışa neden olmaktadır (Çizelge 4.1.1). Şekil 4.1.1’den, 0.90-0.99 planlanan Cronbach alfa değer aralığında, örnek büyüklüğü ortalaması ve güven aralığının diğer üç değer aralığına göre daha düşük seviyelerde olduğu görülmektedir.

Çizelge 4.1.1. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.05 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.05	
	Planlanan Cronbach alfa değer aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	1950,558±284,022
	0,70-0,79	1013,258±204,123
	0,80-0,89	383,274±124,299
	0,90-0,99	60,688±44,860
Madde sayısı=10	0,60-0,69	1733,894±252,468
	0,70-0,79	900,728±181,437
	0,80-0,89	340,733±110,478
	0,90-0,99	54,018±39,855
Madde sayısı=20	0,60-0,69	1642,667±239,183
	0,70-0,79	853,359±171,880
	0,80-0,89	322,832±104,669
	0,90-0,99	51,201±37,755
Madde sayısı=36	0,60-0,69	1605,133±233,708
	0,70-0,79	833,867±167,948
	0,80-0,89	315,460±102,275
	0,90-0,99	50,048±36,908
Madde sayısı=50	0,60-0,69	1592,403±231,854
	0,70-0,79	827,241±166,623
	0,80-0,89	312,964±101,462
	0,90-0,99	49,658±36,609
Madde sayısı=75	0,60-0,69	1581,642±230,292
	0,70-0,79	821,674±165,458
	0,80-0,89	310,850±100,779
	0,90-0,99	49,320±36,350
Madde sayısı=100	0,60-0,69	1576,315±229,515
	0,70-0,79	818,888±164,956
	0,80-0,89	309,807±100,434
	0,90-0,99	49,160±36,233

*0,60-0,69 / 0,70-0,79 ‘orta güvenirlik’
0,80-0,89 / 0,90-0,99 ‘yüksek güvenirlik’

Örnek büyüklüğünün dağılımının Cronbach alfa değer aralığından etkilendiği ancak belirli bir seviyeden sonra ölçekte bulunan madde sayısındaki artıştan etkilenmediği Sekil 4.1.1 ile gösterilmektedir.



Şekil 4.1.1. Paralel ölçüm durumunda, istenilen güven aralığı genişliği 0.05 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

İstenilen güven aralığı genişliğinin 0.1 olduğu senaryo sonuçları Çizelge 4.1.2’de verilmiştir. Planlanan Cronbach alfa değer aralığı düştükçe örnek büyüklüğünün yükseldiği Çizelge 4.1.2’den görülmektedir. Yani, örnek büyüklüğü ortalamasının farklı madde sayılarının alındığı durumların hepsinde planlanan yüksek Cronbach alfa değer aralığında daha düşük olduğunu söylemek mümkündür. Planlanan Cronbach alfa değer aralığının 0.60-0.69’dan 0.90-0.99’a çıkması her farklı madde senaryosunda örnek büyüklüğünde %4’lük bir azalışa neden olmaktadır (Çizelge 4.1.2).

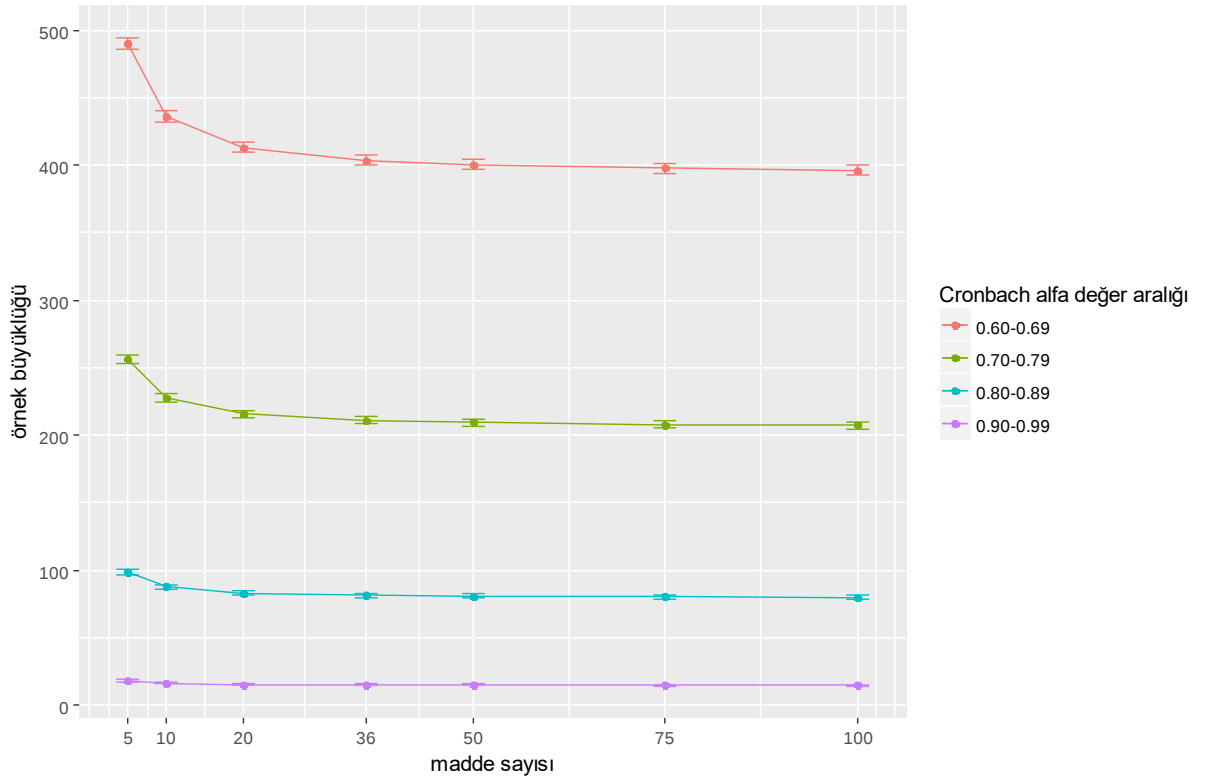
Planlanan Cronbach alfa değer aralıklarında örnek büyüklüğünün ortalaması ve güven aralığına ait grafik Şekil 4.1.2’de verildiği gibidir. Şekil 4.1.2’de 0.90-0.99 planlanan Cronbach alfa değer aralığında örnek büyüklüğünün ortalaması ve güven aralığı diğer üç değer aralığına göre daha düşük seviyelerdedir.

Çizelge 4.1.2. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.1 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.1	
	Planlanan Cronbach alfa değeri aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	490,428±71,004
	0,70-0,79	256,097±51,032
	0,80-0,89	98,619±31,067
	0,90-0,99	18,263±10,983
Madde sayısı=10	0,60-0,69	435,988±63,110
	0,70-0,79	227,695±45,360
	0,80-0,89	87,715±27,613
	0,90-0,99	16,306±9,744
Madde sayısı=20	0,60-0,69	413,078±59,793
	0,70-0,79	215,742±42,975
	0,80-0,89	83,133±26,153
	0,90-0,99	15,470±9,243
Madde sayısı=36	0,60-0,69	403,637±58,426
	0,70-0,79	210,825±41,981
	0,80-0,89	81,247±25,565
	0,90-0,99	15,130±9,042
Madde sayısı=50	0,60-0,69	400,439±57,965
	0,70-0,79	209,158±41,649
	0,80-0,89	80,608±25,360
	0,90-0,99	15,017±8,965
Madde sayısı=75	0,60-0,69	397,742±57,578
	0,70-0,79	207,745±41,376
	0,80-0,89	80,059±25,184
	0,90-0,99	14,923±8,911
Madde sayısı=100	0,60-0,69	396,402±57,379
	0,70-0,79	207,039±41,233
	0,80-0,89	79,794±25,095
	0,90-0,99	14,873±8,878

*0,60-0,69 / 0,70-0,79 ‘orta güvenirlik’
0,80-0,89 / 0,90-0,99 ‘yüksek güvenirlik’

Örnek büyüklüğünün dağılımı ölçekte bulunan madde sayısı arttıkça belirli bir seviyeye kadar azalmakta fakat belirli bir seviyeden sonra ölçekte bulunan madde sayısındaki artıştan etkilenmemektedir. Bu durum, Şekil 4.1.2 ile gösterilmektedir.



Şekil 4.1.2. Paralel ölçüm durumunda istenilen güven aralığı genişliği 0.1 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

İstenilen güven aralığı genişliğinin 0.15 olduğu senaryo sonuçları Çizelge 4.1.3'te verilmiştir. Örnek büyüklüğü ortalamasının, güven aralığı genişliği 0.05 ve 0.1 alındığı durumlarda da olduğu gibi, bütün madde sayıları için planlanan yüksek Cronbach alfa değer aralığında daha düşük olduğunu söylemek mümkündür. Planlanan Cronbach alfa değer aralığının 0.60-0.69'dan 0.90-0.99'a çıkması ise her farklı madde senaryosunda örnek büyüklüğünde %5'lik bir azalışa neden olmaktadır (Çizelge 4.1.3).

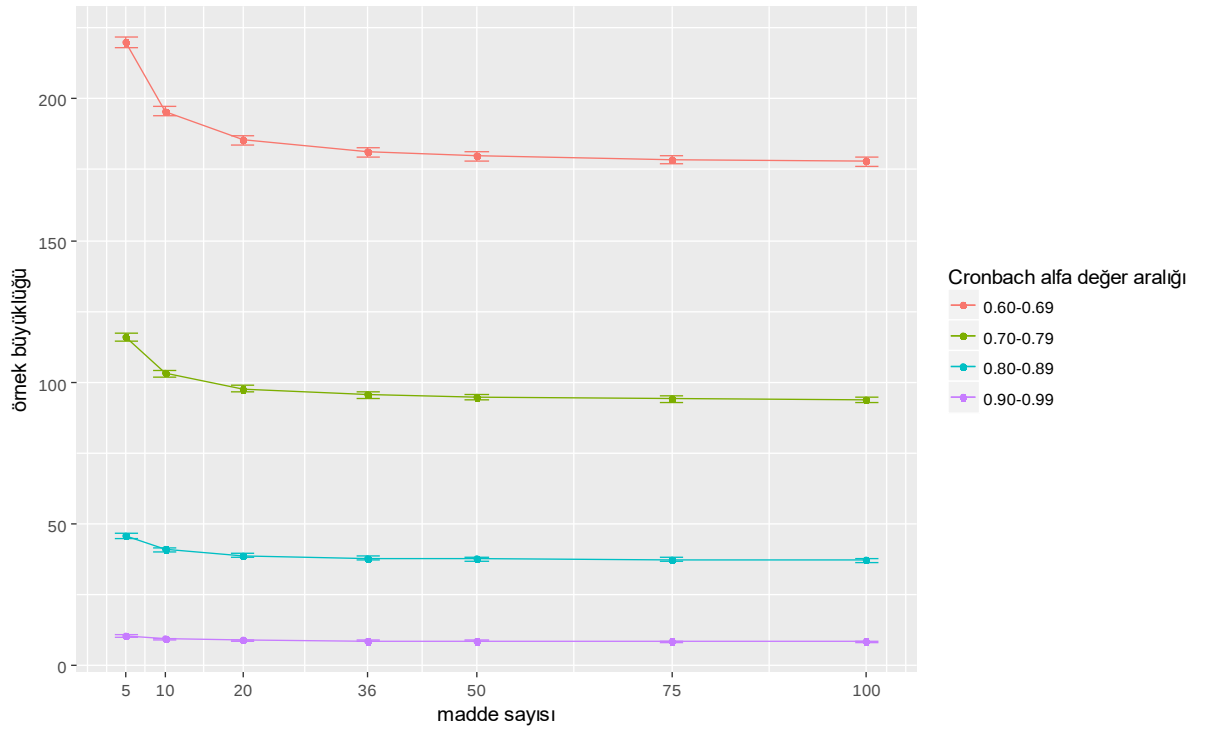
Çizelge 4.1.3. Paralel ölçüm durumunda istenilen güven aralığı genişliğinin 0.15 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.15	
	Planlanan Cronbach alfa değer aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	220,033±31,554
	0,70-0,79	115,897±22,681
	0,80-0,89	45,938±13,794
	0,90-0,99	10,491±4,751
Madde sayısı=10	0,60-0,69	195,635±28,053
	0,70-0,79	103,073±20,161
	0,80-0,89	40,879±12,270
	0,90-0,99	9,413±4,237
Madde sayısı=20	0,60-0,69	185,375±26,571
	0,70-0,79	97,679±19,095
	0,80-0,89	38,765±11,616
	0,90-0,99	8,922±4,015
Madde sayısı=36	0,60-0,69	181,143±25,971
	0,70-0,79	95,462±18,664
	0,80-0,89	37,896±11,352
	0,90-0,99	8,728±3,931
Madde sayısı=50	0,60-0,69	179,712±25,769
	0,70-0,79	94,708±18,510
	0,80-0,89	37,600±11,263
	0,90-0,99	8,664±3,904
Madde sayısı=75	0,60-0,69	178,503±25,583
	0,70-0,79	94,071±18,388
	0,80-0,89	37,351±11,179
	0,90-0,99	8,612±3,878
Madde sayısı=100	0,60-0,69	177,900±25,501
	0,70-0,79	93,754±18,324
	0,80-0,89	37,231±11,150
	0,90-0,99	8,583±3,868

*0,60-0,69 / 0,70-0,79 ‘orta güvenirlik’
0,80-0,89 / 0,90-0,99 ‘yüksek güvenirlik’

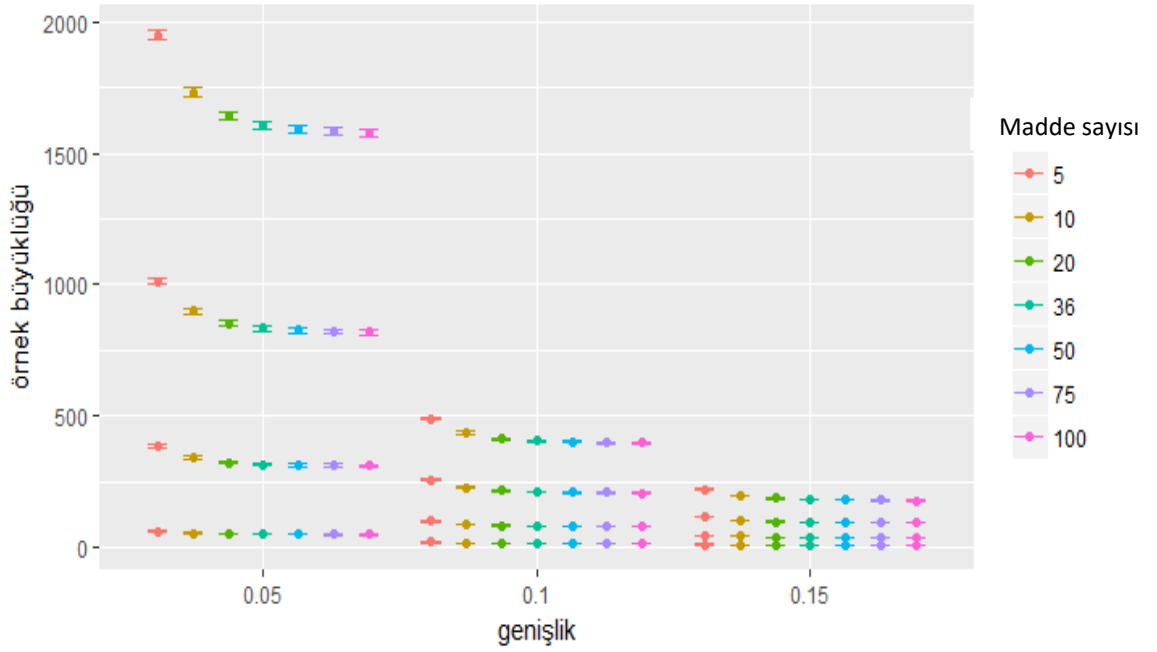
Şekil 4.1.3’te verilen grafik ile istenilen güven aralığı genişliğinin 0.15 olduğu durumda örnek büyüklüğünün ortalaması ve güven aralığı sergilenmektedir. Örnek büyüklüğünün 0.90-0.99 planlanan Cronbach alfa değer aralığında diğer değer aralıklarına göre daha düşük seviyelerde dağıldığı Şekil 4.1.3’ten görülmektedir.

Madde sayısının artmasıyla birlikte bütün Cronbach alfa değer aralıklarında örnek büyüklüğünün önce azaldığı daha sonra stabil hale yaklaştığı Şekil 4.1.3’ten görülmektedir.



Şekil 4.1.3. Paralel ölçüm durumunda istenilen güven aralığı genişliği 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

Uygulaması yapılan istenilen üç güven aralığı genişliğine(0.05, 0.1 ve 0.15) ait toplu hal Şekil 4.1.4'tedir.



†Planlanan Cronbach alfa değer aralığı istenilen üç genişlik durumunda da yukardan aşağıya doğru 0.60-0.69'dan 0.90- 0.99'a doğru artmaktadır.

Şekil 4.1.4. Paralel ölçüm durumunda istenilen güven aralığı genişliği 0.05, 0.1 ve 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

Şekil 4.1.4 de göz önüne alınarak özetlenecek olursa:

- İstenilen güven aralığı genişliğinin 3 durumunda da planlanan Cronbach alfa değer aralığı arttıkça örnek büyüklüğünün azaldığı Şekil 4.1.4'ten görülmektedir.
- İstenilen güven aralığı genişliğinin artmasının örnek büyüklüğünün azalmasına neden olduğu Şekil 4.1.4'ten, istenilen güven aralığı genişliğinin 0.05'ten 0.15'e çıktıkça örnek büyüklüğünün azalmasından görülmektedir.
- Şekil 4.1.4'e göre istenilen üç güven aralığı genişliği durumunda ve bütün planlanan Cronbach alfa değer aralıklarında ölçekte bulunan madde sayısının artmasıyla birlikte örnek büyüklüğü azalmakta ve belirli bir seviyeden sonra azalış durağanlaşmaktadır.

4.1.1.2. Paralel dışı ölçüm varsayımı ile elde edilen simülasyon sonuçları

Paralel dışı ölçüm (tau-eşdeğer) varsayımı altında, istenilen doğruluğa dayalı örnek büyüklüklerine ait simülasyon çalışması sonuçları temel olarak istenen güven aralığı genişliklerinin sırasıyla 0.05, 0.1 ve 0.15 olduğu durumlarda, belirli Cronbach alfa seviyelerinde ve ölçekte bulunan madde sayısına bağlı olarak üretilen varyans-kovaryans matrisleri ele alınarak incelenmiştir.

İstenilen güven aralığı genişliğinin 0.05 olduğu senaryo sonuçları Çizelge 4.1.4’de verilmiştir. Aynı planlanan Cronbach alfa değer aralıklarında madde sayısının artmasıyla elde edilen örnek büyüklüğünün azaldığı görülmektedir. Bununla beraber, madde sayısının spesifik olduğu durumlarda (Ör, madde sayısı=5), planlanan Cronbach alfa değer aralığı seviyesi arttıkça örnek büyüklüğü azalmaktadır(Çizelge 4.1.4).

İstenilen güven aralığı genişliğinin 0.1 olduğu senaryo sonuçları Çizelge 4.1.5’te ve istenilen güven aralığı genişliğinin 0.15 olduğu senaryo sonuçları ise Çizelge 4.1.6’da verilmiştir. Güven aralığı genişliğinin 0.05 olduğu senaryodaki örnek büyüklüğüne dair değerlendirmelerin, genişliğin 0.1 ve 0.15 olduğu durumlara da genellenebileceği Çizelge 4.1.5 ve Çizelge 4.1.6’dan anlaşılmaktadır.

Çizelge 4.1.4. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.05 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.05	
	Planlanan Cronbach alfa değer aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	1857,387±308,955
	0,70-0,79	939,354±219,085
	0,80-0,89	339,253±124,531
	0,90-0,99	142,120±11,667
Madde sayısı=10	0,60-0,69	1673,102±272,973
	0,70-0,79	847,116±194,866
	0,80-0,89	288,122±112,886
	0,90-0,99	80,662±27,071
Madde sayısı=20	0,60-0,69	1577,529±255,978
	0,70-0,79	794,976±183,505
	0,80-0,89	280,996±108,311
	0,90-0,99	44,846±31,148
Madde sayısı=36	0,60-0,69	1542,865±249,387
	0,70-0,79	778,577±179,761
	0,80-0,89	264,452±103,618
	0,90-0,99	28,872±29,933
Madde sayısı=50	0,60-0,69	1525,291±253,517
	0,70-0,79	759,071±178,550
	0,80-0,89	265,930±103,806
	0,90-0,99	23,872±28,971
Madde sayısı=75	0,60-0,69	1514,158±250,544
	0,70-0,79	757,268±180,430
	0,80-0,89	263,768±106,359
	0,90-0,99	18,914±26,460
Madde sayısı=100	0,60-0,69	1394,801±164,715
	0,70-0,79	761,570±172,111
	0,80-0,89	263,796±103,946
	0,90-0,99	15,835±23,630

*0,60-0,69 / 0,70-0,79 ‘orta güvenirlik’

0,80-0,89 / 0,90-0,99 ‘yüksek güvenirlik’

Çizelge 4.1.5. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.1 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.1	
	Planlanan Cronbach alfa değer aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	465,602±77,239
	0,70-0,79	236,118±54,752
	0,80-0,89	86,092±31,164
	0,90-0,99	36,769±8,937
Madde sayısı=10	0,60-0,69	419,278±68,258
	0,70-0,79	212,798±48,721
	0,80-0,89	73,032±28,220
	0,90-0,99	21,214±6,736
Madde sayısı=20	0,60-0,69	395,271±63,997
	0,70-0,79	199,640±45,897
	0,80-0,89	71,177±27,084
	0,90-0,99	12,275±7,702
Madde sayısı=36	0,60-0,69	386,583±62,349
	0,70-0,79	195,500±44,954
	0,80-0,89	66,978±25,896
	0,90-0,99	8,395±7,337
Madde sayısı=50	0,60-0,69	382,181±63,385
	0,70-0,79	190,604±44,646
	0,80-0,89	67,347±25,951
	0,90-0,99	7,319±6,996
Madde sayısı=75	0,60-0,69	379,360±62,630
	0,70-0,79	190,163±45,090
	0,80-0,89	66,773±26,566
	0,90-0,99	6,311±6,278
Madde sayısı=100	0,60-0,69	349,515±41,164
	0,70-0,79	191,220±43,019
	0,80-0,89	66,800±25,965
	0,90-0,99	5,652±5,547

*0,60-0,69 / 0,70-0,79 ‘orta güvenirlik’

0,80-0,89 / 0,90-0,99 ‘yüksek güvenirlik’

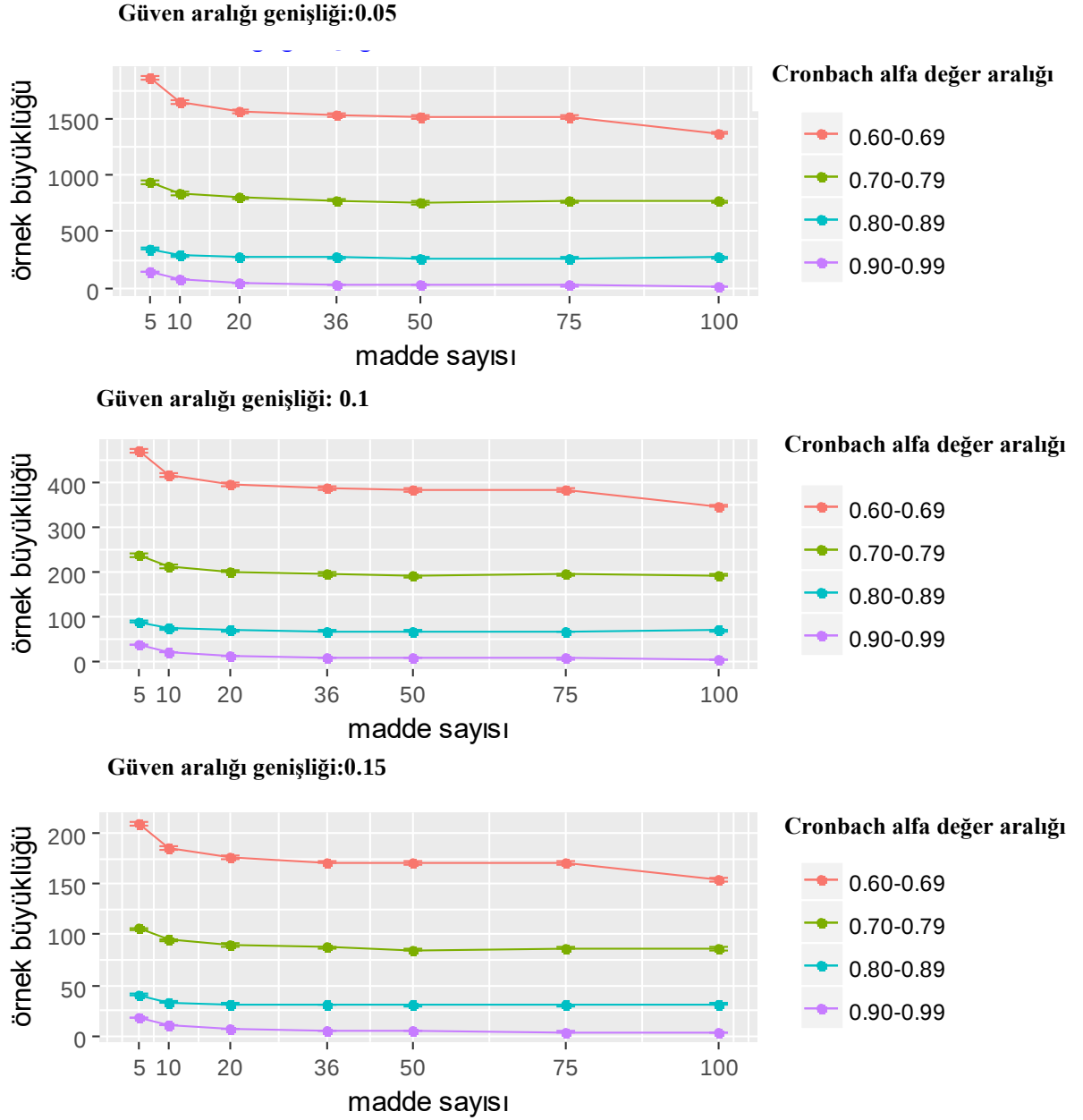
Çizelge 4.1.6. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliğinin 0.15 alınması ile elde edilen örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Genişlik=0.15	
	Planlanan Cronbach alfa değer aralığı*	Örnek Büyüklüğü ort±ss
Madde sayısı=5	0,60-0,69	207,889±34,330
	0,70-0,79	105,882±24,346
	0,80-0,89	39,161±13,846
	0,90-0,99	17,248±1,339
Madde sayısı=10	0,60-0,69	187,092±30,337
	0,70-0,79	95,323±21,663
	0,80-0,89	33,242±12,524
	0,90-0,99	10,264±2,981
Madde sayısı=20	0,60-0,69	176,352±28,432
	0,70-0,79	89,396±20,395
	0,80-0,89	32,326±12,039
	0,90-0,99	6,333±3,366
Madde sayısı=36	0,60-0,69	172,448±27,707
	0,70-0,79	87,514±19,959
	0,80-0,89	30,439±11,486
	0,90-0,99	4,876±3,030
Madde sayısı=50	0,60-0,69	170,476±28,167
	0,70-0,79	85,335±19,838
	0,80-0,89	30,579±11,518
	0,90-0,99	4,518±2,842
Madde sayısı=75	0,60-0,69	169,212±27,849
	0,70-0,79	85,139±20,033
	0,80-0,89	30,324±11,800
	0,90-0,99	4,173±2,519
Madde sayısı=100	0,60-0,69	155,943±18,294
	0,70-0,79	85,613±19,117
	0,80-0,89	30,320±11,531
	0,90-0,99	3,908±2,183

*0,60-0,69 / 0,70-0,79 ‘orta güvenilirlik’

0,80-0,89 / 0,90-0,99 ‘yüksek güvenilirlik’

Paralel dışı ölçüm durumunda, istenilen farklı güven aralığı genişliklerinde örnek büyüklüklerinin Cronbach alfa değer aralıkları ve ölçekte bulunan madde sayıları ile ilişkisi Şekil 4.1.5'te verilmiştir.



Şekil 4.1.5. Paralel dışı ölçüm durumunda istenilen güven aralığı genişliği 0.05, 0.1 ve 0.15 iken örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

Güven aralığı genişliğinin 0.05'ten 0.15'e çıkmasının örnek büyüklüğünde yüksek bir azalışa neden olduğu Şekil 4.1.5'ten görülmektedir. İstenilen üç güven aralığı genişliği durumunda da, Cronbach alfa değer aralığı arttıkça ve/veya ölçekte bulunan madde sayısı arttıkça örnek büyüklüğünün azaldığı Şekil 4.1.5'ten görülmektedir.

4.1.2. İstenilen Güce Dayalı Örnek Büyüklüğü için Simülasyon Sonuçları

Burada yapılan simülasyon çalışması, ' $H_0: \alpha = h$ ' hipotezini 0.80, 0.85 veya 0.90 güç ile test etmek için gerekli örnek büyüklüğü nedir?' sorusu üzerine kuruludur. Bu sorunun cevabı, ölçekte bulunan farklı madde sayıları ve planlanan farklı Cronbach alfa değer aralıkları dikkate alınarak değerlendirilmiştir. Değerlendirmelere dair sonuçlar Çizelge 4.1.7 ile verilmiştir. Sonuçların görsel olarak değerlendirilebilmesi adına Şekil 4.1.6 verilmiştir.

Planlanan Cronbach alfa değer aralığı 0.60-0.69 iken güç değerinin bütün durumları için madde sayısı arttıkça örnek büyüklüğü azalmakta ancak belirli bir seviyeden sonra azalış durağanlaşmaktadır(Çizelge 4.1.7). Bu durumun, aynı zamanda planlanan tüm Cronbach alfa değer aralıkları için geçerli olduğu Şekil 4.1.6'dan görülmektedir.

Çizelge 4.1.7. $H_0: \alpha = h$ hipotezini test etmek için 0.80, 0.85 ve 0.90 güç değerlerine, farklı madde sayılarına ve planlanan Cronbach alfa değer aralıklarına dayalı örnek büyüklüğüne ait tanımlayıcılar

SENARYOLAR	Madde sayısı	Planlanan Cronbach alfa değer aralığı 0.60-0.69	Planlanan Cronbach alfa değer aralığı 0.70-0.79	Planlanan Cronbach alfa değer aralığı 0.80-0.89	Planlanan Cronbach alfa değer aralığı 0.90-0.99
		Örnek Büyüklüğü $ort \pm ss$	Örnek Büyüklüğü $ort \pm ss$	Örnek Büyüklüğü $ort \pm ss$	Örnek Büyüklüğü $ort \pm ss$
Güç=0.80	5	181,54±86,635	106,721±53,233	51,582±27,894	15,937±10,728
	10	161,632±77,005	95,142±47,317	46,125±24,798	14,443±9,536
	20	153,257±72,952	90,264±44,830	43,828±23,496	13,814±9,038
	36	149,819±71,293	88,261±43,805	42,887±22,964	13,553±8,827
	50	148,648±70,735	87,578±43,462	42,565±22,776	13,466±8,758
	75	147,657±70,240	87,007±43,161	42,298±22,620	13,392±8,702
	100	147,172±70,008	86,712±43,013	42,160±22,540	13,354±8,668
Güç=0.85	5	207,288±99,090	121,724±60,897	58,640±31,912	17,865±12,270
	10	184,533±88,084	108,470±54,121	52,403±28,366	16,162±10,910
	20	174,957±83,445	102,906±51,297	49,776±26,873	15,440±10,332
	36	171,012±81,539	100,593±50,103	48,698±26,269	15,143±10,097
	50	169,670±80,898	99,825±49,706	48,329±26,046	15,042±10,014
	75	168,547±80,357	99,161±49,370	48,023±25,875	14,957±9,945
	100	167,987±80,087	98,843±49,218	47,868±25,784	14,918±9,917
Güç=0.90	5	242,173±115,975	142,020±71,259	68,206±37,353	20,481±14,354
	10	215,539±103,081	126,523±63,354	60,902±33,204	18,484±12,758
	20	204,329±97,675	119,999±60,018	57,833±31,452	17,646±12,094
	36	199,713±95,441	117,302±58,642	56,564±30,731	17,295±11,809
	50	198,148±94,675	116,397±58,182	56,135±30,488	17,182±11,723
	75	196,829±94,040	115,626±57,792	55,778±30,285	17,085±11,644
	100	196,174±93,720	115,246±57,585	55,587±30,175	17,034±11,599

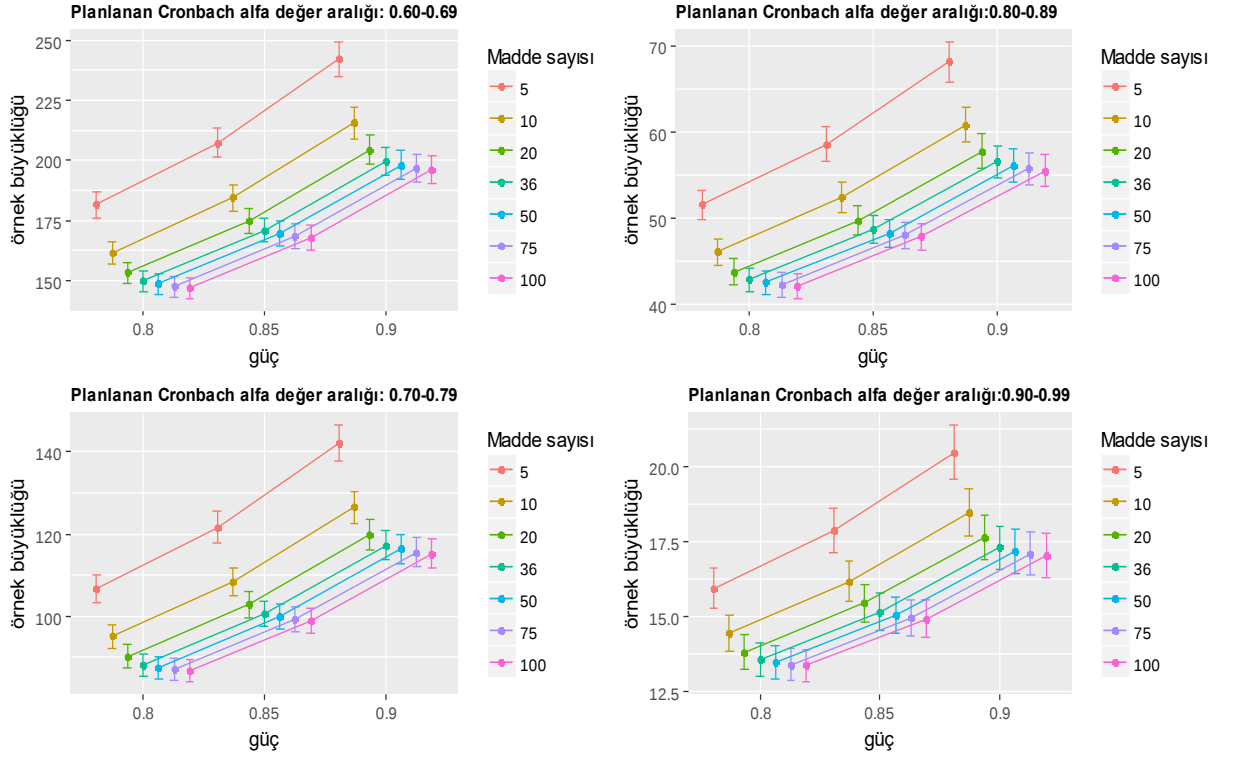
*Planlanan Cronbach alfa değer aralığı 0.60-0.69 iken; $h=0.50$

Planlanan Cronbach alfa değer aralığı 0.70-0.79 iken; $h=0.60$

Planlanan Cronbach alfa değer aralığı 0.80-0.89 iken; $h=0.70$

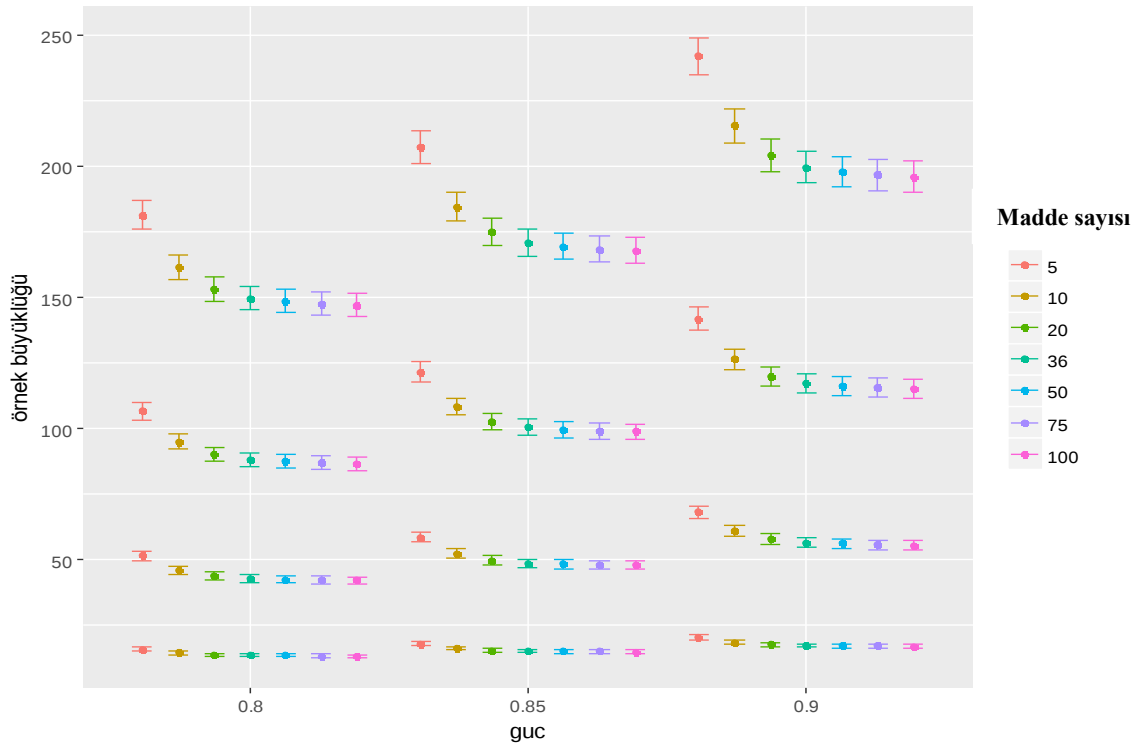
Planlanan Cronbach alfa değer aralığı 0.90-0.99 iken; $h=0.80$

Ölçekte bulunan madde sayısının aynı alındığı durumda planlanan Cronbach alfa değer aralığı 0.60-0.69'dan 0.90-0.99'a arttıkça güç değerinin tüm durumları için örnek büyüklüğünde azalış söz konusudur(Çizelge 4.1.7). Örneğin, güç=0.90 ve madde sayısı=5 iken planlanan Cronbach alfa değer aralığının artması ile örnek büyüklüğünde yaşanan düşüş Şekil 4.1.6 ile açık bir şekilde ifade edilmiştir.



Şekil 4.1.6. Planlanan farklı Cronbach alfa değer aralıklarında $H_0: \alpha = h$ hipotezini test etmek için gerekli örnek büyüklüğü ortalamasının ve ortalamanın güven aralığının madde sayısına ve güce bağlı dağılımı

Planlanan bütün Cronbach alfa değer aralıklarında aynı madde sayıları göz önüne alındığında, güç değerinin 0.80'den 0.90'a çıkması örnek büyüklüğünde artışa sebep olmaktadır(Çizelge 4.1.7). Şekil 4.1.6'dan da görüldüğü gibi tüm planlanan Cronbach alfa değer aralıkları için güç değeri arttıkça örnek büyüklüğü daha yüksek seviyelerdedir.



†Planlanan Cronbach alfa değer aralığı istenilen üç güç durumunda da yukardan aşağıya doğru 0.60-0.69'dan 0.90- 0.99'a doğru artmaktadır.

Şekil 4.1.7. İstenilen güç değeri 0.8, 0.85 ve 0.9 iken örnek büyüklüğü ortalamasının ve ortalamamanın güven aralığının madde sayısına ve Cronbach alfa değer aralığına bağlı dağılımı

Uygulaması yapılan istenilen üç güç değerine ait toplu hal Şekil 4.1.7 ile sergilenmiştir. Bu grafik göz önüne alınarak özetlenecek olursa:

- Şekil 4.1.7'de bütün istenilen güç değerleri için 0.90-0.99 planlanan Cronbach alfa değer aralığında örnek büyüklüğünün ortalaması ve güven aralığı diğer değer aralıklarına göre daha düşük seviyelerdedir. Bu durumu sırasıyla, 0.80-0.89, 0.70-0.79 ve 0.60-0.69 Cronbach alfa değer aralıkları izlemektedir.
- Şekil 4.1.7'den görüldüğü gibi bütün istenilen güç değerleri için örnek büyüklüğünün dağılımı ölçekte bulunan madde sayısı arttıkça belirli bir seviyeye kadar azalmakta ancak belirli bir seviyeden sonra ölçekte bulunan madde sayısındaki artıştan etkilenmemektedir.
- İstenilen güç değeri arttıkça Cronbach alfa değer aralıkları ve madde sayılarının tüm durumları için örnek büyüklüğünün de arttığı Şekil 4.1.7'den görülmektedir.

4.2. Gerçek Veri Setine ait Bulgular

4.2.1. İstenilen Doğruluğa Dayalı Örnek Büyüklüğü Değerlendirilmesi

4.2.1.1. Paralel Ölçüm Durumunda Örnek Büyüklüğünün Değerlendirilmesi (ODI)

Gerçek veri seti uygulaması kapsamında ele alınan ODI verisinin, ölçüm durumlarından hangisine uyduğunun değerlendirilebilmesi adına, veriden rastgele 1000 tekrarlı 20 örnek çekilerek oluşturulan varyans ve kovaryans değerlerine ait güven aralıklarını içeren 10×10 boyutlu matris Çizelge 4.2.1'de verilmiştir. Matrisin köşegeninde bulunan güven aralıklarının birbirlerini içerdiği Çizelge 4.2.1'den görülmektedir. Bu durum, madde varyanslarının birbirine yakın olduğunun bir göstergesidir. Ek olarak, matrisin köşegen dışı elemanlarında bulunan güven aralıklarının da birbirlerini içerdiği Çizelge 4.2.1'den görülmektedir. Yani, maddeler arası kovaryanslar da benzerdir. Sonuç olarak, varyans ve kovaryansların benzer olması paralel ölçüm durumuna karşılık gelmektedir. Bu nedenle, ölçeğin Cronbach alfa katsayısı için yapılan güven aralığı hesaplamalarında paralel ölçüme dayalı yaklaşım kullanılmıştır.

10 maddeden oluşan ODI ölçeğinden rastgele örnekleme ile $n=20, 40, 80, 120, 160$ ve 200 örnek çekilerek elde edilen Cronbach alfa değerlerine ait tanımlayıcılar Çizelge 4.2.2'de ve güven aralığı bilgisi de göz önünde alınarak grafiksel gösterim Şekil 4.2.1'de verilmiştir.

Şekil 4.2.1'de, farklı örnek büyüklükleri ile yapılan çekimlerde elde edilen Cronbach alfa değerlerinin ortalamaları ve güven aralıkları yer almaktadır. Örnek büyüklüğü arttıkça Cronbach alfa değeri ortalaması ve güven aralığının daha yüksek seviyelere yerleştiği Şekil 4.2.1'den görülmektedir. Ancak, Cronbach alfa değeri ortalamasının, örnek büyüklüğü belirli bir seviyeye geldikten sonra durağanlaştığı ve bu örnek büyüklüğü seviyesinin $n=120$ olduğu Çizelge 4.2.2'den görülmektedir.

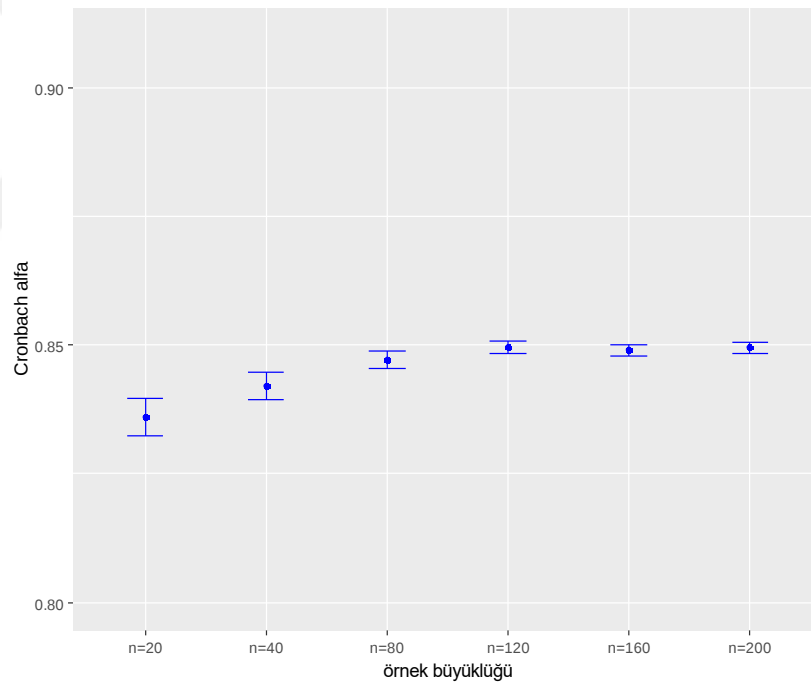
Örnek büyüklüğünün artmasıyla Cronbach alfa'ya bağlı güven aralığı genişliğinin azaldığı da Şekil 4.2.1'den görülmektedir.

Çizelge 4.2.1.ODI'nin, elemanları %95'lik güven aralığından oluşan varyans-kovaryans matrisi

	Ağrının Şiddeti	Kişisel Bakım	Yük Kaldırma	Yürüme	Oturma	Ayakta Durma	Uyuma	Cinsel Yaşam	Sosyal Yaşam	Seyahat
Ağrının Şiddeti	(0,228;1,123)									
Kişisel Bakım	(-0,068;0,638)	(0,192;1,197)								
Yük Kaldırma	(-0,185;0,797)	(-0,089;0,920)	(0,737;2,493)							
Yürüme	(-0,031;0,592)	(-0,019;0,799)	(-0,144;0,928)	(0,169;1,505)						
Oturma	(-0,134;0,771)	(-0,253;0,675)	(-0,243;1,091)	(-0,173;0,739)	(0,334;1,916)					
Ayakta Durma	(-0,158;0,608)	(-0,205;0,686)	(-0,007;1,044)	(-0,036;0,711)	(0,035;0,927)	(0,329;1,559)				
Uyuma	(-0,072;0,531)	(-0,143;0,661)	(-0,222;0,955)	(-0,090;0,618)	(-0,153;0,768)	(-0,065;0,797)	(-0,131;1,626)			
Cinsel Yaşam	(-0,137;1,791)	(-0,168;0,636)	(-0,336;0,920)	(-0,104;0,595)	(-0,174;0,672)	(-0,120;0,692)	(-0,454;1,242)	(-0,236;1,791)		
Sosyal Yaşam	(-0,093;0,600)	(-0,145;0,614)	(-0,149;1,111)	(-0,099;0,615)	(-0,061;0,794)	(-0,038;0,746)	(-0,231;0,932)	(-0,288;0,925)	(0,074;1,309)	
Seyahat	(-0,091;0,609)	(-0,106;0,547)	(-0,148;0,926)	(-0,026;0,724)	(-0,065;0,911)	(0,033;0,673)	(-0,109;0,671)	(-0,099;0,648)	(-0,067;0,697)	(-0,066;1,289)

Çizelge 4.2.2.ODI'den rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler

ODI	Cronbach Alfa Ort±Ss
n=20	0,836±0,059
n=40	0,842±0,042
n=80	0,847±0,026
n=120	0,849±0,019
n=160	0,849±0,017
n=200	0,849±0,015



Şekil 4.2.1. ODI'de Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

Paralel ölçümlere dayalı simülasyon çalışmasına bakıldığında, Cronbach alfa değeri aralığı 0.80-0.90 ve güven aralığı genişliği $\cong 0.1$ iken ODI ölçeği için gerekli örnek büyüklüğü $87,715 \pm 27,613$ 'tür. Burada, tutucu olmak adına üst limit alınırsa, gerekli örnek büyüklüğü 116'dır denilebilir. Bu durumu, Çizelge 4.2.2'den de görüldüğü gibi örnek büyüklüğü 120 olduğunda Cronbach alfa değeri ortalamasının durağanlaşması ile de destekleyebiliriz.



4.2.1.2. Paralel Dışı Ölçüm Durumunda Örnek Büyüklüğünün Değerlendirilmesi (SF-36)

Gerçek veri seti uygulaması kapsamında ele alınan bir diğer veri SF-36 ölçeğine aittir. SF-36 ölçeğinin alt boyutlarının kendi aralarında anlamlı (* $p < 0.001$) ve pozitif yönde ilişkili olduğu Çizelge 4.2.3'ten görülmektedir.

Çizelge 4.2.3. SF-36 alt ölçeklerine ait korelasyon analizi

	FF	FRG	ERG	VA	RS	V	Sİ	GSA
FF	-	0,624*	0,623*	0,654*	0,450*	0,645*	0,458*	0,571*
FRG		-	0,702*	0,617*	0,412*	0,598*	0,533*	0,598*
ERG			-	0,586*	0,502*	0,624*	0,527*	0,609*
VA				-	0,496*	0,678*	0,628*	0,602*
RS					-	0,714*	0,508*	0,592*
V						-	0,574*	0,691*
Sİ							-	0,464*
GSA								-

Örnek büyüklüğünü değerlendirme kapsamında, SF-36 ölçeğinin alt boyutlarının ölçüm durumlarından hangisine uyduğunun belirlenmesi için FF alt ölçeği baz alınmış ve bu ölçeğe dair bulgulara detaylı bir şekilde yer verilmiştir.

Çizelge 4.2.4.FF’in, elemanları %95’lik güven aralığından oluşan varyans-kovaryans matrisi

	Madde 1	Madde 2	Madde 3	Madde 4	Madde 5	Madde 6	Madde 7	Madde 8	Madde 9	Madde 10
Madde 1	(585,71;1990,84)									
Madde 2	(556,06;1761,95)	(1464,94;2536,62)								
Madde 3	(256,92;1637,06)	(394,67;1843,130)	(864,25;2328,80)							
Madde 4	(238,24;1475,49)	(488,48;1758,34)	(430,87;1724,04)	(846,79;2084,88)						
Madde 5	(42,91;956,22)	(81,41;1400,78)	(137,13;1153,71)	(159,13;1178,24)	(557,45;1909,12)					
Madde 6	(119,81;1361,94)	(173,38;1719,10)	(198,80;1620,60)	(242,12;1613,22)	(21,05;1329,28)	(1222,31;2364,30)				
Madde 7	(432,25;1732,41)	(531,33;1995,32)	(402,76;1930,35)	(431,24;1876,18)	(92,66;1238,02)	(448,27;1913,32)	(1237,36;2536,19)			
Madde 8	(500,50;1595,12)	(719,52;2154,55)	(371,55;1717,24)	(554,83;1808,89)	(228,12;1500,85)	(399,08;1842,30)	(700,12;2032,50)	(1464,88;2548,31)		
Madde 9	(152,49;997,08)	(133,784;1372,58)	(67,22;1084,88)	(208,73;1224,27)	(-41,23;1246,03)	(-77,39;1220,88)	(84,36;1226,08)	(470,54;1749,48)	(423,60;1987,97)	
Madde 10	(43,15;894,80)	(118,162;1320,64)	(120,01;1032,40)	(139,03;1255,20)	(-117,51;1255,20)	(9,34;1164,42)	(104,94;1147,09)	(145,18;1462,48)	(-38,14;1390,75)	(314,42;1884,07)

FF ölçeğine ait verinin, ölçüm durumlarından hangisine uyduğunun değerlendirilebilmesi adına, veriden rastgele 1000 tekrarlı 20 örnek çekilerek oluşturulan varyans ve kovaryans değerlerine ait güven aralıklarını içeren 10×10 boyutlu matris Çizelge 4.2.4'te verilmiştir. Matrisin köşegen ve köşegen dışı elemanları incelendiğinde güven aralıklarının birbirini içermediği Çizelge 4.2.4'den görülmektedir. Bu durum, madde varyanslarının ve maddeler arası kovaryansların benzer olmadığının bir göstergesidir. Sonuç olarak, varyans ve kovaryansların benzer olmaması paralel dışı ölçüm(konjenerik) durumuna karşılık gelmektedir. Bu nedenle, ölçeğin Cronbach alfa katsayısı için yapılan güven aralığı hesaplamalarında paralel dışı ölçüme dayalı yaklaşım kullanılmıştır.

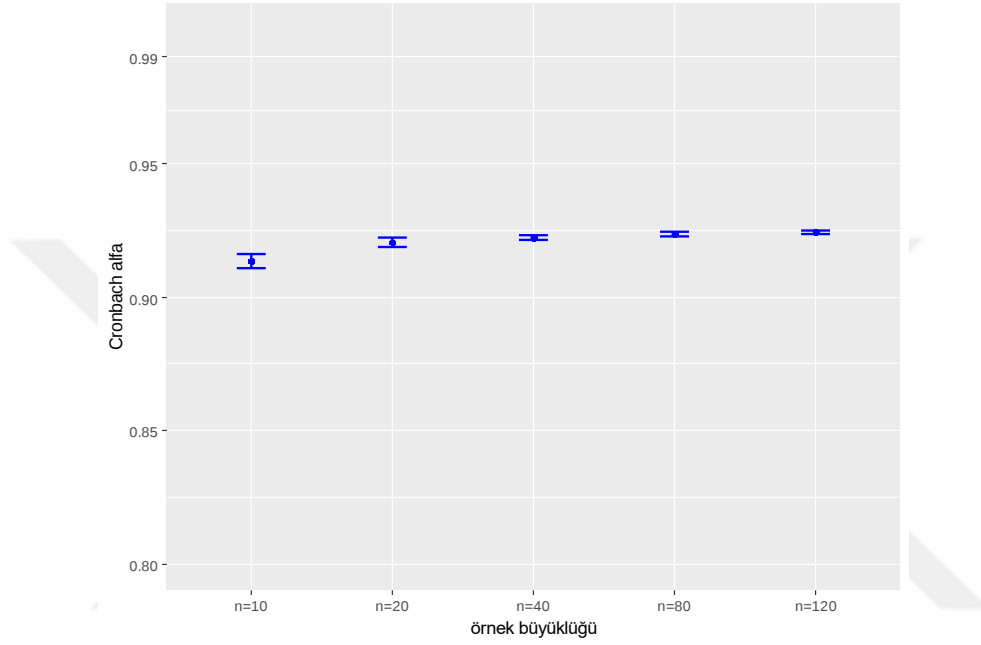
4.2.1.2.1. Örnek Büyüklüğü Çalışması Sonuçları

10 maddeden oluşan FF ölçeğinden rastgele örnekleme ile $n=10, 20, 40, 80$ ve 120 örnek çekilerek elde edilen Cronbach alfa değerlerine ait tanımlayıcılar Çizelge 4.2.5'te verilmiş ve güven aralığı bilgisi de göz önüne alınarak Şekil 4.2.2 ile desteklenmiştir.

Çizelge 4.2.5.FF ölçeğinden rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler

FF	Cronbach Alfa Ort±Ss
n=10	0,913±0,045
n=20	0,920±0,025
n=40	0,922±0,016
n=80	0,923±0,011
n=120	0,924±0,008

Örnek büyüklüğü arttıkça Cronbach alfa için daha durağan tahminler elde edildiği Çizelge 4.2.5'ten görülmektedir. Şekil 4.2.2, örnek büyüklüğü arttıkça Cronbach alfa değerlerinin daha yüksek seviyelere yerleştiğini bunun yanı sıra, güven aralığı genişliğinin azaldığını göstermektedir



Şekil 4.2.2 FF ölçeğinde Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

SF-36 yaşam kalitesi ölçeğinin geriye kalan 7 alt boyutunun (FRG, ERG, V, VA, GSA, Sİ, RS) ölçüm durumlarını değerlendirmek adına varyans-kovaryans matrisleri incelenmiş ve FF ölçeği ile aynı, yani, paralel dışı ölçüm(konjenerik) durumuna denk düştükleri görülmüştür. Bu yüzden, bu alt ölçeklerin Cronbach alfa katsayısı için yapılan güven aralığı hesaplamalarında da paralel dışı ölçüme dayalı yaklaşım kullanılmıştır.

SF-36 ölçeğinin alt boyutlarına ait, rasgele örnekleme ile n=10, 20, 40, 80 ve 120 örnek çekilerek elde edilen Cronbach alfa değerlerine dair tanımlayıcılar Çizelge 4.2.6'da verilmiştir. Çizelge 4.2.6 incelendiğinde, örnek büyüklüğü arttıkça bütün alt ölçeklerde Cronbach alfa değer ortalamasının arttığı görülmektedir.

Şekil 4.2.3'te alt ölçeklerin farklı örnek büyüklüğü ile yapılan çekimlerde elde edilen Cronbach alfa değer ortalamaları ve güven aralıkları verilmiştir. Bütün alt ölçeklerde, örnek büyüklüğü arttıkça Cronbach alfanın daha yüksek seviyelerde dağıldığı ve Cronbach alfaya bağlı güven aralığı genişliğinin azaldığı Şekil 4.2.3 ile görsel olarak ifade edilmiştir.

Şekil 4.2.4 ile alt ölçeklerin örnek büyüklüğüne bağlı Cronbach alfa değeri ortalamaları ve güven aralıkları toplu olarak verilmiştir. Eldeki veriden hesaplanan RS ve V alt ölçeklerinin Cronbach alfa değeri ortalamalarının diğer alt ölçeklere göre daha düşük seviyelerde dağıldığı, FRG alt ölçeğinin Cronbach alfa değeri ortalamasının ise diğer alt ölçeklere göre daha yüksek seviyelerde dağıldığı Şekil 4.2.4'ten görülmektedir. Yine Şekil 4.2.4'ten, bütün alt ölçeklerde hesaplanan Cronbach alfa değerinin örnek büyüklüğü arttıkça durağanlaştığı gözlemlenmektedir.

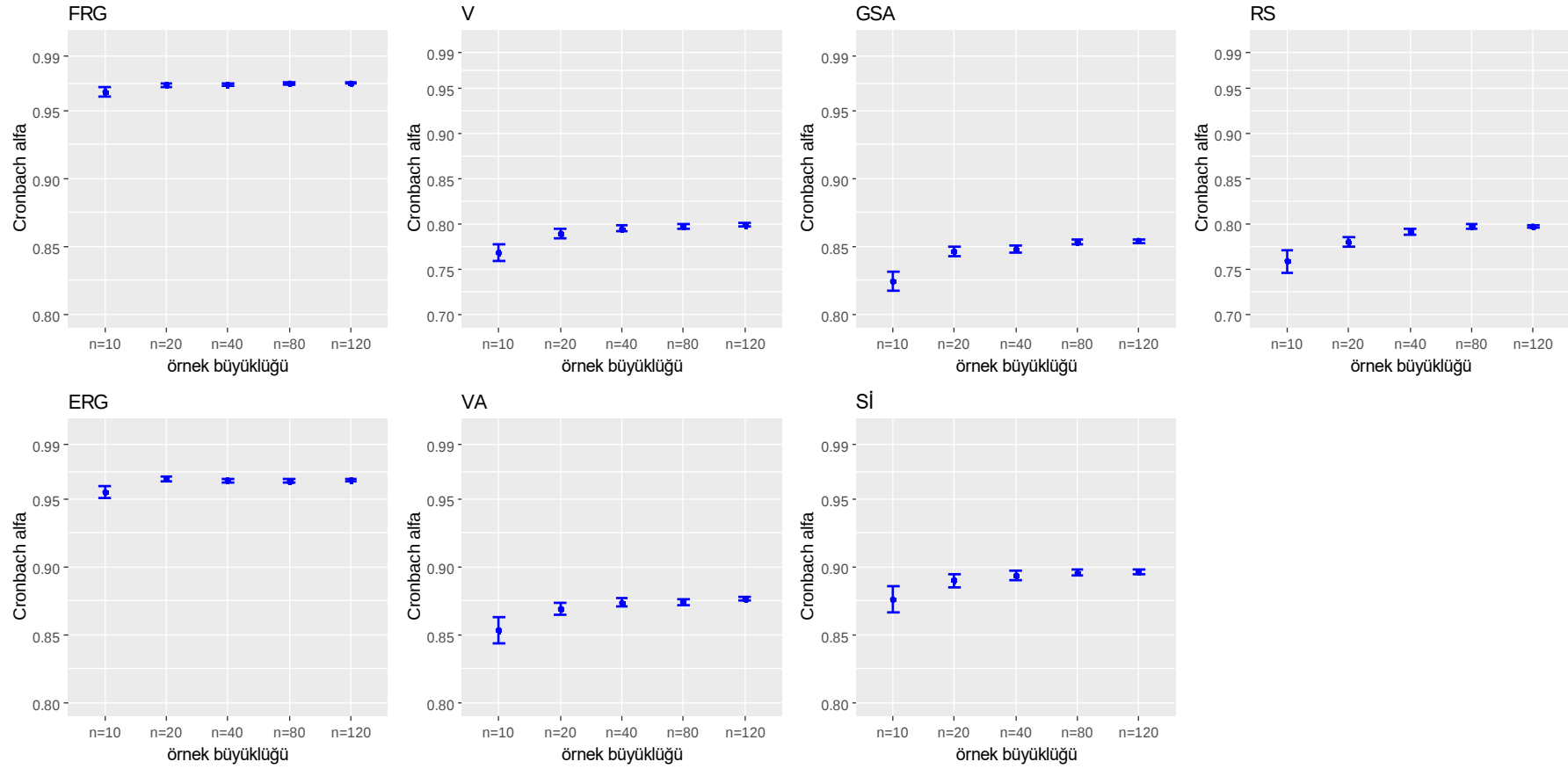
Çizelge 4.2.6.SF-36 alt ölçeklerinin (FRG, ERG, V, VA, GSA, Sİ, RS) rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler

FRG		ERG	
Cronbach Alfa		Cronbach Alfa	
Ort±Ss		Ort±Ss	
n=10	0,963±0,058	n=10	0,961±0,051
n=20	0,969±0,023	n=20	0,963±0,030
n=40	0,969±0,016	n=40	0,964±0,020
n=80	0,969±0,011	n=80	0,965±0,014
n=120	0,970±0,009	n=120	0,965±0,011

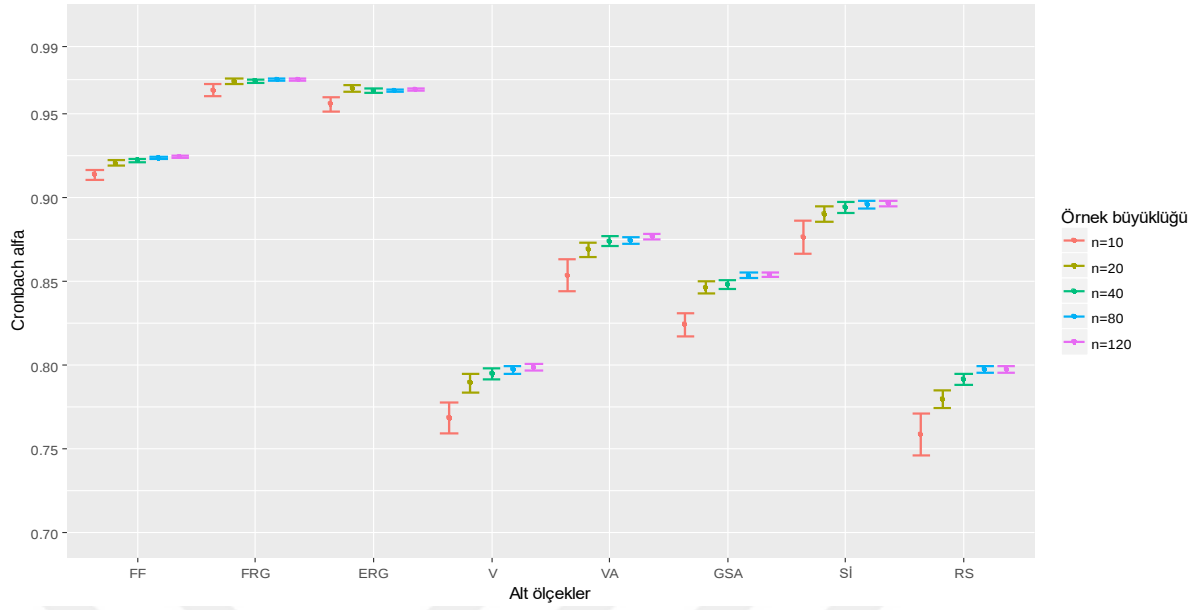
V		VA	
Cronbach Alfa		Cronbach Alfa	
Ort±Ss		Ort±Ss	
n=10	0,768±0,147	n=10	0,853±0,151
n=20	0,789±0,090	n=20	0,868±0,070
n=40	0,794±0,054	n=40	0,873±0,047
n=80	0,797±0,038	n=80	0,874±0,033
n=120	0,798±0,030	n=120	0,876±0,026

GSA		Sİ	
Cronbach Alfa		Cronbach Alfa	
Ort±Ss		Ort±Ss	
n=10	0,824±0,113	n=10	0,876±0,157
n=20	0,846±0,058	n=20	0,889±0,072
n=40	0,848±0,038	n=40	0,893±0,051
n=80	0,853±0,024	n=80	0,895±0,034
n=120	0,854±0,020	n=120	0,896±0,028

RS	
Cronbach Alfa	
Ort±Ss	
n=10	0,758±0,197
n=20	0,779±0,086
n=40	0,791±0,053
n=80	0,797±0,035
n=120	0,797±0,029



Şekil 4.2.3. SF-36 alt ölçeklerinde Cronbach alfa değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı



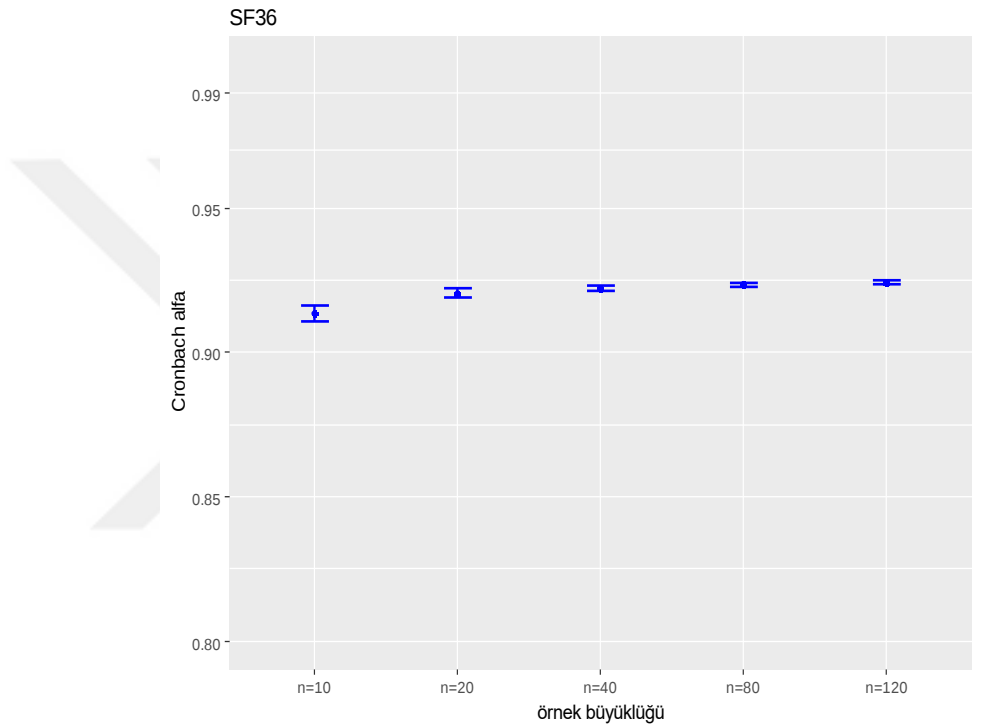
Şekil 4.2.4. SF-36 alt ölçeklerinin Cronbach alfa değerleri ortalaması ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

Paralel dışı ölçüm durumunda örnek büyüklüğünü değerlendirmek adına SF-36 ölçeği bir bütün halinde de ele alınmıştır. SF-36 ölçeğine ait rasgele örnekleme ile n=10, 20, 40, 80 ve 120 örnek çekilerek elde edilen Cronbach alfa değerlerine ait tanımlayıcılara Çizelge 4.2.7’de yer verilmiştir. SF-36 ölçeğinde farklı örnek büyüklüğü ile yapılan çekimlerde elde edilen Cronbach alfa değer ortalamaları ve güven aralıkları Şekil 4.2.5 ile gösterilmiştir.

Çizelge 4.2.7. SF-36 ölçeğinin rastgele örnekleme ile elde edilen Cronbach alfa değerleri için tanımlayıcı istatistikler

SF-36	Cronbach Alfa Ort±Ss
n=10	0,955±0,022
n=20	0,959±0,013
n=40	0,960±0,007
n=80	0,961±0,005
n=120	0,961±0,004

Çizelge 4.2.7’den örnek büyüklüğündeki artışın Cronbach alfa değeri üzerindeki etkisinin minimal olduğu saptanmıştır. Örnek büyüklüğü arttıkça Cronbach alfa değerlerinin daha yüksek seviyelerde dağıldığı ancak belirli bir seviyeden sonra durağan hale geldiği Şekil 4.2.5’ten görülmektedir.



Şekil 4.2.5. SF-36 ölçeğinin Cronbach alfa değerleri ortalaması ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

4.2.1.2.2. Güven Aralığı Çalışması Sonuçları

Paralel dışı ölçümler için simülasyon çalışması sonuçları ile yapılacak olan değerlendirmelerde, ilk olarak SF-36 ölçeğinin alt boyutları baz alınmıştır. Alt boyutlardan FRG, ERG, V, VA ve Sİ’ nin madde sayıları 5’ten küçük olduğundan değerlendirme FF, GSA ve RS alt boyutları ile yapılmıştır. Simülasyon çalışmasında FF, GSA ve RS alt boyutlarına denk düşen örnek büyüklüğü sınıflarına Çizelge 4.2.8’de yer verilmiştir.

Çizelge 4.2.8.SF-36 ölçeğinin FF-GSA-RS alt boyutlarına paralel dışı simülasyon çalışmasında denk gelen örnek büyüklüğü sınıfı

Alt Boyut	Güven Aralığı Genişliği	Örnek Büyüklüğü
FF	$\cong 0.05$	80,662 $\pm$ 27,071
GSA	$\cong 0.1$	86,092 $\pm$ 31,164
RS	$\cong 0.15$	39,161 $\pm$ 13,846

Ölçek için hesaplanacak Cronbach alfa değerinin doğruluğu adına, örnek büyüklüğü değerinde tutucu olunması yani, değerlendirmeye dahil edilen FF, GSA ve RS ölçekleri için simülasyon çalışmasında elde edilen örnek büyüklüğü değerlerinden en büyük olanının üst limitinin alınması tercih edilir. Bu sebeple, simülasyon çalışmasına bağlı olarak alt ölçekler bazında örnek büyüklüğü değeri 118 olarak tespit edilir.

İkinci olarak, paralel dışı ölçümler için simülasyon çalışması sonuçları ile yapılacak olan değerlendirmelerde, SF-36 ölçeği bir bütün halinde ele alınmıştır. Simülasyon çalışmasında SF-36 ölçeğine denk düşen örnek büyüklüğü sınıfı, güven aralığı genişliğinin yaklaşık olarak 0.05 alınması ile 28,872 $\pm$ 29,933 olarak elde edilmiştir. Tutucu olmak adına üst limit ele alındığında gerekli örnek büyüklüğü 59 olarak elde edilir.

4.2.2. İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirilmesi

4.2.2.1.ODI Verisinde İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirmesi

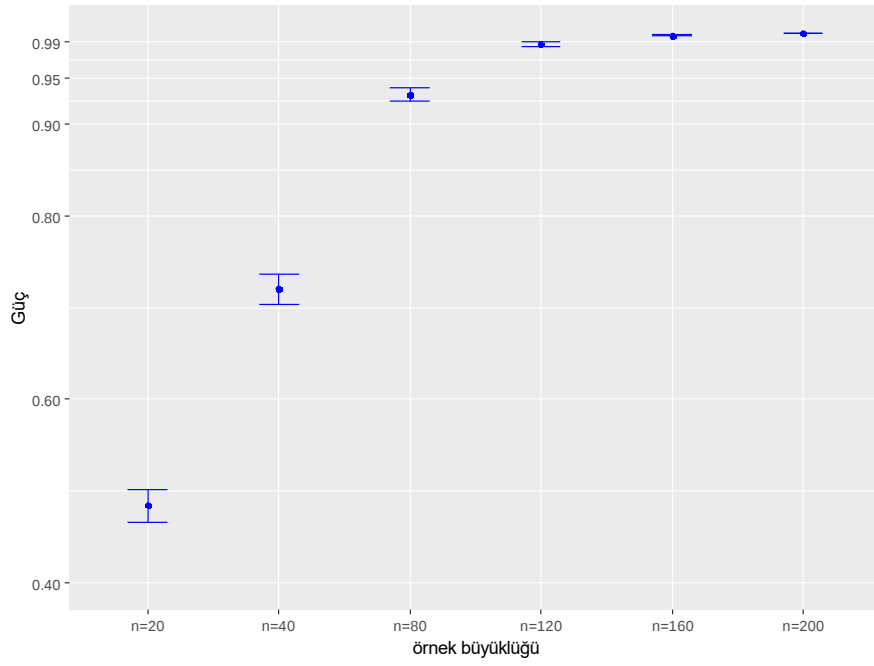
ODI verisine ait istenilen doğruluk için yapılan çalışmada elde edilen Çizelge 4.2.2'den çeşitli örnek büyüklüklerinde Cronbach alfa değerinin 0.830-0.850 değer aralığında değiştiği görülmüştür. Dolayısıyla, bu bölümde $H_0: \alpha = 0.70$ hipotezi baz alınarak yapılan değerlendirilmelere yer verilmiştir.

ODI ölçeğinden rastgele örnekleme ile $n=20, 40, 80, 120, 160$ ve 200 örnek çekilerek elde edilen güç değeri için tanımlayıcılar Çizelge 4.2.9'da yer verilmiş ve güven aralığı bilgisi de göz önüne alınarak Şekil 4.2.6 ile desteklenmiştir.

Çizelge 4.2.9.ODI ölçeğinin rastgele örnekleme ile elde edilen güç değeri için tanımlayıcı istatistikler

ODI	Güç Ort±Ss
n=20	0,483±0,284
n=40	0,720±0,259
n=80	0,932±0,118
n=120	0,987±0,043
n=160	0,997±0,012
n=200	0,999±0,002

Düşük örnek büyüklüklerinde $H_0: \alpha = 0.70$ hipotezini test etmedeki gücün de düşük olduğu ancak örnek büyüklüğü arttıkça $H_0: \alpha = 0.70$ hipotezini test etmedeki gücün arttığı hatta $n=200$ durumunda 1'e çok yakın olduğu görülmektedir(Çizelge 4.2.9). Şekil 4.2.6'dan da görüldüğü gibi, güç değerlerinin ortalamasının ve güven aralığının yüksek seviyelerde dağıldığı durumda, örnek büyüklüğü de yüksek seviyelerdedir.



Şekil 4.2.6 ODI’de güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

İstenilen güç için yapılan simülasyon çalışmasında istenilen güç seviyesi olarak %80, %85 veya %90’a ulaşmak adına ODI ölçeğinin kaç bireye uygulanması gerektiği sorusunun cevabı Çizelge 4.2.10’dan görülmektedir.

Çizelge 4.2.10.ODI’nin, istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı

Güç Değeri	Örnek Büyüklüğü
%80	46,125±24,798
%85	52,403±28,366
%90	60,902±33,240

Çizelge 4.2.10'dan 10 maddeden oluşan ODI ölçeğinde, istenilen gücün %90 olması durumunda tutucu olursa bile gerekli örnek büyüklüğünün 95 olduğu, ancak gerçek veri setinin ölçeğin 212 kişiye uygulanarak elde edildiği bilinmektedir. Bu durum, güç değerinin %99'lara kadar çıkmasını sağlamıştır. Simülasyon çalışmasında elde edilen örnek büyüklüğü değeri (95), $H_0: \alpha = 0.70$ hipotezini %90 güç ile test etmemizi sağlamak için yeterlidir sonucuna varılabilir.

4.2.2.1. SF-36 Verisinde İstenilen Güce Dayalı Örnek Büyüklüğü Değerlendirmesi

4.2.2.1.1. SF-36 Alt Ölçekleri ile Güç Değerlendirmesi

İlk olarak SF-36 alt ölçekleri baz alınarak güç değerlendirmeleri yapılmıştır. SF-36 verisine ait elde edilen Cronbach alfa değerleri alt ölçeklerde çeşitlilik göstermektedir. Bu sebeple, $H_0: \alpha = h$ hipotezinde yer alan h değeri her alt ölçek için farklıdır ve Çizelge 4.2.11'de verilmiştir.

Çizelge 4.2.11.SF-36 alt ölçeklerinin $H_0: \alpha = h$ hipotezindeki h değerleri

Alt Ölçekler	Cronbach Alfa	h
FF	0.924	0.80
FRG	0.970	0.80
ERG	0.965	0.80
V	0.800	0.70
VA	0.876	0.70
GSA	0.855	0.70
SI	0.897	0.70
RS	0.799	0.70

SF-36 alt ölçeklerinden rasgele örnekleme ile $n=10, 20, 40, 80$ ve 120 örnek çekilmesi sonucu elde edilen güç değerlerine ait tanımlayıcılara Çizelge 4.2.12'de, güç değerlerin ortalaması ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımına ise Şekil 4.2.7'de yer verilmiştir.

FF	Güç Ort±Ss
n=10	0,456±0,257
n=20	0,731±0,229
n=40	0,939±0,104
n=80	0,997±0,020
n=120	0,999±0,0008

FRG	Güç Ort±Ss
n=10	0,795±0,02
n=20	0,956±0,01
n=40	0,997±0,001
n=80	0,999±5×10 ⁻⁵
n=120	0,999±4×10 ⁻¹¹

ERG	Güç Ort±Ss
n=10	0,726±0,031
n=20	0,898±0,018
n=40	0,990±0,004
n=80	0,999±5×10 ⁻⁵
n=120	0,999±3×10 ⁻⁶

V	Güç Ort±Ss
n=10	0,218±0,208
n=20	0,296±0,249
n=40	0,391±0,271
n=80	0,568±0,287
n=120	0,711±0,256

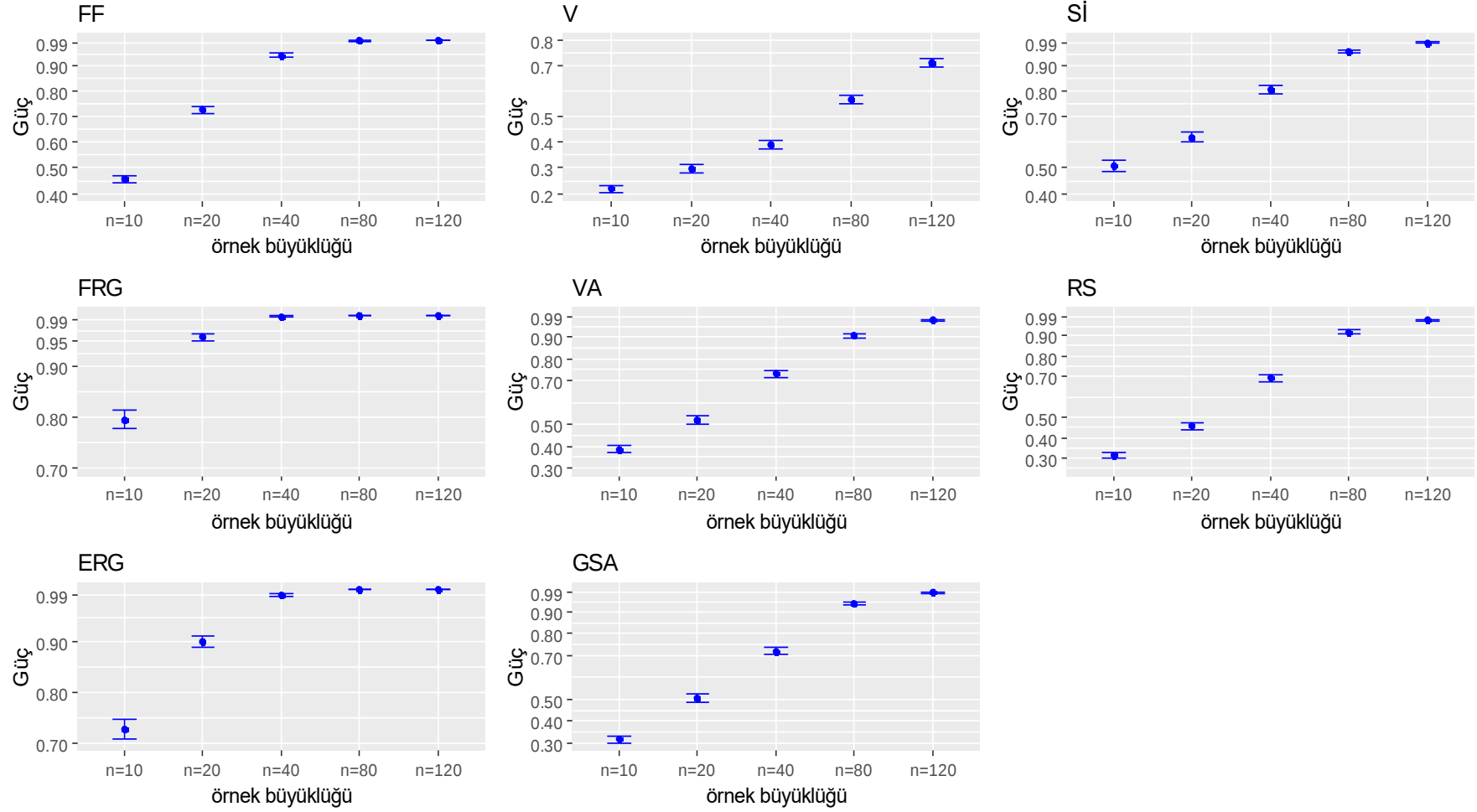
VA	Güç Ort±Ss
n=10	0,387±0,298
n=20	0,520±0,309
n=40	0,730±0,263
n=80	0,907±0,152
n=120	0,975±0,075

GSA	Güç Ort±Ss
n=10	0,316±0,249
n=20	0,505±0,286
n=40	0,719±0,250
n=80	0,940±0,100
n=120	0,986±0,042

Si	Güç Ort±Ss
n=10	0,507±0,347
n=20	0,620±0,329
n=40	0,806±0,255
n=80	0,955±0,110
n=120	0,989±0,039

RS	Güç Ort±Ss
n=10	0,314±0,250
n=20	0,456±0,273
n=40	0,691±0,258
n=80	0,917±0,132
n=120	0,975±0,059

Çizelge 4.2.12. SF-36 alt ölçeklerinin rastgele örnekleme ile elde edilen güç değerleri için tanımlayıcı istatistikler



Şekil 4.2.7. SF-36 alt ölçeklerinde güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

Alt ölçekler genel olarak ele alındığında, örnek büyüklüğü arttıkça sıfır hipotezini test etmedeki gücün de arttığı Çizelge 4.2.12’den görülmektedir. Bu durumu Şekil 4.2.7, örnek büyüklüğü arttıkça güç değeri ortalamasının ve güven aralığının daha yüksek seviyelerde dağılması ile desteklenmektedir.

Paralel dışı ölçüm durumunda örnek büyüklüğünün değerlendirilmesi kapsamında SF-36 ele alındığında madde sayılarının az olması sebebiyle bazı alt ölçekler değerlendirmeye dahil edilmemiş, değerlendirmeler FF, GSA ve RS alt boyutları ile yapılmıştır. Aynı ölçekler ele alınarak, istenilen güç için yapılan simülasyon çalışmasında istenilen güç seviyesi olarak %80, %85 veya %90’a ulaşmak adına alt boyutların kaç bireye uygulanması gerektiği sorusunun cevabı Çizelge 4.2.13’ten görülmektedir.

Çizelge 4.2.13. SF-36 alt ölçeklerinden FF, GSA ve RS’in istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı

Güç Değeri	Örnek Büyüklüğü	
	FF	GSA-RS
%80	14,443±9,536	51,582±27,894
%85	16,162±10,910	58,640±31,912
%90	18,484±12,758	68,206±37,353

Alt ölçekler baz alındığında istenilen gücün %90 olması durumunda gerekli örnek büyüklüğünün tutucu olmak adına ölçekler arasında maksimum örnek büyüklüğüne sahip olanın alınması uygun bulunmuştur. Çizelge 4.2.13’ten görüldüğü gibi, uygun örnek büyüklüğünün 106 olduğu ancak gerçek veri setinin de ölçeğin 120 kişiye uygulanarak elde edildiği bilinmektedir.

4.2.2.1.2. Bir Bütün Olarak SF-36 ile Güç Değerlendirmesi

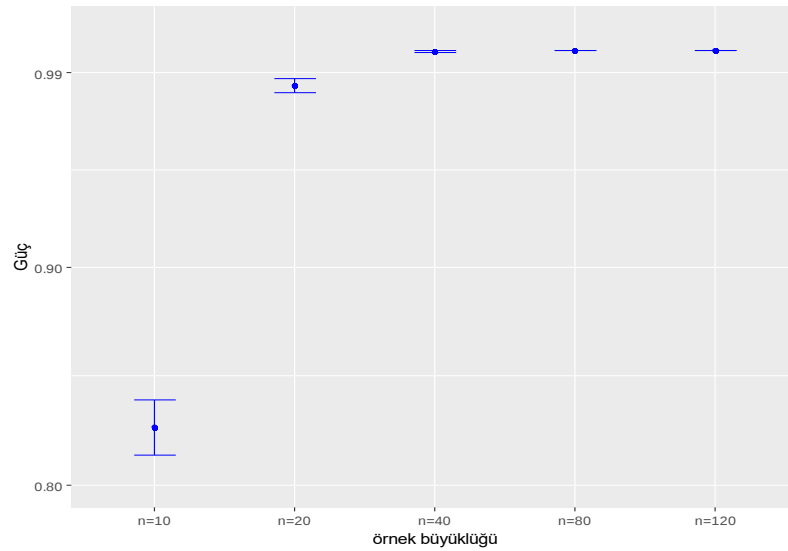
SF-36 ölçeği bir bütün olarak ele alındığında istenilen doğruluk için yapılan çalışmada elde edilen Cronbach değeri (0.961) göz önünde bulundurularak, $H_0: \alpha = h$

hipotezinde yer alan h değeri 0.80 alınmış ve gerekli güç hesaplamalarına ait tanımlayıcı istatistiklere Çizelge 4.2.14’te yer verilmiştir.

Çizelge 4.2.14. SF-36 ölçeğinin rastgele örnekleme ile elde edilen güç değeri için tanımlayıcı istatistikler

SF-36	Güç Ort±Ss
n=10	0.826±0.206
n=20	0,983±0,005
n=40	0,999±0,0008
n=80	0,999±2×10 ⁻⁹
n=120	0,999±3×10 ⁻¹⁶

$H_0: \alpha = 0.80$ hipotezini test etmede, düşük örnek büyüklüklerinde gücün de düşük olduğu ancak, örnek büyüklüğü arttıkça hipotezi test etmedeki gücün de arttığı Çizelge 4.2.14’ten görülmektedir. Bu durumu Şekil 4.2.8, örnek büyüklüğü arttıkça güç değeri ortalamasının ve güven aralığının daha yüksek seviyelerde dağılması ile desteklemektedir.



Şekil 4.2.8. SF-36 ölçeğinde güç değerinin ortalamasının ve ortalamanın güven aralığının örnek büyüklüğüne bağlı dağılımı

İstenilen güç için yapılan simülasyon çalışmasında istenilen güç seviyesi olarak %80, %85 veya %90'a ulaşmak adına SF-36 ölçeğinin kaç bireye uygulanması gerektiği sorusunun cevabı Çizelge 4.2.15 ile verilmektedir.

Çizelge 4.2.15.SF-36 ölçeğinin, istenilen güç için örnek büyüklüğü simülasyon çalışmasına göre denk düştüğü örnek büyüklüğü sınıfı

Güç Değeri	Örnek Büyüklüğü
%80	13,553±8,827
%85	15,143±10,097
%90	17,295±11,809

36 maddeden oluşan SF-36 ölçeğinde, istenilen gücün %90 olması durumunda gerekli örnek büyüklüğünün 30 olduğu, ancak gerçek veri setinin ölçeğin 120 kişiye uygulanarak elde edildiği bilinmektedir. Bu durum, güç değerinin %99'lara kadar çıkmasını sağlamıştır. Simülasyon çalışmasında elde edilen örnek büyüklüğü değeri (30), $H_0: \alpha = 0.80$ hipotezini %90 güç ile test etmemizi sağlamak için yeterlidir sonucuna varılabilir.

5. TARTIŞMA

Geçerlik ve güvenirlik, sağlık alanında kullanılan ölçeklerin taşınması gereken vazgeçilmez iki ana özelliktir. Bu tez çalışması kapsamında detaylı olarak ele alınan güvenirlik, klasik test teorisinde genel olarak, gerçek skorların varyansının hata skorlarının varyansına oranı olarak elde edilir. Fakat gerçek skorların varlığı hiçbir zaman bilinmediğinden bu durum hipotetiktir ve güvenirliği dolaylı olarak hesaplamak adına farklı yöntemler geliştirilmiştir¹.

Sağlık alanında en sık kullanılan güvenirlik ölçütü Cronbach'ın alfasıdır. En genel uygulaması olan ölçümlerin çoklu ölçek maddelerini temsil ettiği durumlarda Cronbach alfa, iç tutarlılık güvenirliğinin ölçüsü olarak tanımlanır⁸. Literatürde Cronbach alfa'nın, ölçüm modellerine bağlı olarak farklı tanımlamaları mevcuttur.

- Alfa katsayısı paralel ya da tau-eşdeğer ölçüm modellerindeki güvenirliğe eşittir⁴².
- Alfa katsayısı, tau-eşdeğer ölçüm durumundaki güvenirliğe eşittir^{11,29,32}.

Bu ölçüm modelleri, klasik test teorisine dayanır ve bütün maddelerin varyansları, kovaryansları, korelasyonları birbirine eşitse paralel model, madde varyansları eşit değil ancak maddeler arası kovaryanslar eşit ise tau-eşdeğer model söz konusudur. Bu modellere ek olarak konjenerik model de eşit olmayan madde varyansları ve kovaryansları varlığında kullanılabilir⁴⁵. Konjenerik modellerde genellikle, alfa güvenirliği olduğundan az tahmin eder. Sijtsma ve Green&Yang, konjenerik ölçüm olması durumunda alfa katsayısını kullanmak yerine, alternatif bir yaklaşım kullanarak güvenirliğe ilişkin daha iyi bir tahmin elde edilebileceğini ileri sürmüşlerdir.^{46,47}

Literatürde, sonlu örneklemelerin, madde varyansları ve kovaryansları yaklaşık olarak birbirine eşit olduğunda araştırmacıları paralel modele yönlendirdiği ve Likert tipi ölçeklerin paralel ölçüm modeline uyduğu yönünde varsayımlar söz konusudur⁴². Bunlara ek olarak; paralel ölçüm modelinin çok kısıtlayıcı olduğu söylenmekte, buna rağmen literatürde Cronbach'ın alfasına bağlı güven aralığı yaklaşımlarının paralel modele dayandığı görülmektedir^{48,49}. Cronbach'ın alfasının tau-eşdeğer ölçüm durumunda optimal sonuç verdiğini söyleyen Bonett(2015), paralel dışı ölçüm durumlarına dayalı

güven aralığı yaklaşımı önermiş, bu güven aralığı yaklaşımına bağlı ve $H_0: \alpha = h$ hipotezini istenilen güç ile test etmeye bağlı olmak üzere iki örnek büyüklüğü yöntemi geliştirmiştir⁸.

Bu çalışmayla, önerilen güven aralığı ve örnek büyüklüğü yaklaşımlarına dayanarak, istenilen doğruluk için yapılan, paralel ve paralel dışı ölçüm varsayımlarına dayalı simülasyonlarda örnek büyüklüğünün, madde sayısından, istenilen güven aralığı genişliğinden ve Cronbach alfa değer aralığından etkilendiği sonucuna ulaşılmıştır. Buna ek olarak, istenilen doğruluğa dayalı simülasyon çalışmasında, aynı madde sayıları, Cronbach alfa değer aralıkları ve istenilen güven aralığı genişliğinin alınması durumunda paralel ölçüme dayalı olarak hesaplanan örnek büyüklüğü değerlerinin, paralel dışı ölçüme dayalı hesaplanan örnek büyüklüğü değerlerinden biraz daha yüksek olduğu görülmüştür. Bunun sebebi, paralel ölçüm yaklaşımının daha kısıtlayıcı olması olarak düşünülebilir.

İstenilen doğruluğa dayalı simülasyon çalışması sonuçları ile literatüre paralel olarak aşağıdaki bilgilere ulaşılmıştır:

- İstenilen güven aralığı genişliğinin aynı olduğu durumlarda, yüksek alfa değerleri daha az örnek büyüklüğü gerektirmiştir⁴. Yani, madde sayısı sabit olduğu durumda planlanan yüksek güvenirlilik düzeyleri daha düşük örnek büyüklükleri elde edilmesine neden olur.
 - Örneğin, paralel ölçüm durumunda, madde sayısı 10 ve güven aralığı genişliği 0.1 iken, 0.70-0.79 alfa değer aralığı 274, 0.80-0.89 alfa değer aralığı ise 116 örnek büyüklüğü gerektirmektedir.
 - Paralel dışı ölçüm durumunda örnek verecek olursak, madde sayısı 10 ve güven aralığı genişliği 0.1 iken, 0.70-0.79 alfa değer aralığı 262, 0.80-0.89 alfa değer aralığı ise 102 örnek büyüklüğü gerektirmektedir.
- İstenilen güven aralığı daraldıkça, örnek büyüklüğü daha yüksek seviyelerde dağılmıştır^{9,10,39}.
 - İstenilen güven aralığı genişliği 0.15'ten 0.05'e düştükçe örnek büyüklüğü değerlerinde artış gözlemlenmiştir.

- Ölçekteki madde sayısı arttıkça (10'a kadar), gerekli örnek büyüklüğü hızla düşmektedir. Ancak daha sonra düşüşteki artış azalmaktadır⁴.
 - Hem paralel hem de paralel dışı ölçüm durumunda, alınan spesifik güven aralığı genişliklerinde (0.05, 0.1 veya 0.15), orta ve yüksek güvenirlilik düzeylerinde madde sayısı 50 olduğunda örnek büyüklüğü sabitlenmektedir.

İstenilen güç için yapılan simülasyon çalışması sonuçlarından, bütün istenilen güç değerlerinde (0.80, 0.85 ve 0.90), yüksek Cronbach alfa değer aralıklarında örnek büyüklüğünün düştüğü görülmektedir. Bunun yanısıra, literatüre paralel olarak sıfır hipotezini test etmedeki güç arttıkça, çalışmalar için gerekli örnek büyüklüğünün de arttığı gözlemlenmiştir³⁹.

İstenilen güç için yapılan simülasyon çalışmasından, istenilen doğruluk için yapılan simülasyon çalışmasında olduğu gibi madde sayısının artması örnek büyüklüğünü belirli bir seviyeye kadar düşürmekte belirli bir seviyeden sonra ise azalış durağanlaşmaktadır. Planlanan Cronbach alfa değer aralıklarının hepsinde, en geç madde sayısı 50'den sonra örnek büyüklüğü sabitlenmektedir.

Gerçek veri çalışmasında ele alınan ODI ölçek verisinin, ölçüm modellerinden hangisine uyduğunu belirlemek adına varyans ve kovaryans değerlerinin güven aralığını içeren bir matris oluşturulmuş ve paralel ölçüm modeline uyduğu saptanmıştır. İstenilen doğruluk için yapılan simülasyon çalışmasına göre alınması gereken örnek büyüklüğü sayısı 116 iken $H_0: \alpha = 0.70$ hipotezini %90 güç ile test etmek adına alınması gerekli örnek büyüklüğü 94 olarak bulunmuştur. Her iki durumda elde edilen örnek büyüklüğü de ODI ölçeğinin eldeki gerçek uygulamasında alınan 212'den küçük olduğu görülmektedir. Kullanılan ODI çalışmasının, gerektiğinden fazla kişiye uygulanarak yapıldığı sonucuna ulaşılır.

Gerçek veri çalışmasında ele alınan SF-36 ölçek verisinin, varyans-kovaryans değerlerinin güven aralığından oluşan matrisin incelenmesi ile paralel dışı ölçüm(konjenerik) durumuna uyduğu gözlemlenmiştir. Bu kapsamda, istenilen doğruluk için yapılan simülasyon çalışmasına göre alınması gereken örnek büyüklüğü, alt boyutlar baz alındığında 118 iken, ölçek bir bütün olarak ele alındığında 59'dur. Bu hesaplamada

kullanılan simülasyon çalışmasında üretilmiş olan varyans-kovaryans matrisleri tau-eşdeğer modele uyduğundan, SF-36'nın ise konjenerik ölçüme uyduğu tahmin edildiğinden, elde edilen örnek büyüklüğü düşük olabilir.

SF-36 ele alınarak istenilen güç için yapılan simülasyon çalışmasında ise örnek büyüklüğü, alt boyutlar baz alındığında 106 iken, ölçek bir bütün olarak ele alındığında 30'dur.

Literatürde, 0.90'ın üzerinde alfa katsayısı değerlerinin fazlalıkları temsil ettiğini ve ölçüm aracının kısaltılması gerektiğini belirten yayınlar mevcuttur¹⁹. SF-36 bir bütün olarak ele alındığında Cronbach alfa değeri 0.961 olarak bulunmuştur. SF-36'da hem istenilen doğruluk için hem de istenilen güç için alt boyutlar baz alındığında ve ölçek bir bütün olarak ele alındığında elde edilen iki örnek büyüklüğü arasındaki farklılık, Cronbach alfa değerinin yüksekliğinden ve alt boyutlara sahip olan bir ölçekte Cronbach alfa değerinin alt boyutlar baz alınarak hesaplanması gerektiğinden kaynaklanmaktadır. Bu kapsamda, eldeki SF-36 çalışmasının 120 kişiye uygulanarak yapıldığı göz önüne alınırsa, alt boyutlar baz alınarak elde edilen istenilen doğruluğa bağlı örnek büyüklüğü 118 ve istenilen güce bağlı örnek büyüklüğü 106 olduğundan, optimal sonuçların elde edilmesi için örnek büyüklüğünün yeterli olduğu sonucuna ulaşılır.

ODI ve SF-36 ölçek verilerinden rasgele örnek çekilerek elde edilen Cronbach alfa değerlerinde, örnek büyüklüğü arttıkça artış gözlemlenmekte ancak artış miktarının minimal seviyelerde olduğu dikkat çekmektedir. Buradan, literatüre paralel olarak Cronbach alfa tahminleri için çok büyük örneklerle gerek olmadığı, küçük örneklerle bile popülasyon parametresinin temsil edebileceği sonucuna varılmıştır⁵⁰.

Sonuç olarak, herhangi bir bilimsel araştırmada olduğu gibi güvenilirlik çalışması yaparken de esas adım, çalışmanın amacında verilene denk düşen minimum örnek büyüklüğünün hesaplanmasıdır. Literatürde örnek büyüklüğü ile ilgili farklı yaklaşımlar mevcuttur:

- Nunnally, likert tipi ölçekleri kullanırken araştırmacılara en az 300 katılımcıyı örnek büyüklüklerine dahil etmelerini önerir³⁶.
- Charter, güvenilirlik için daha kesin tahminler elde etmek adına örnek büyüklüğünün 400 alınmasını önerir^{6,9}.

- Javali, iyi bir iç tutarlılık tahmini elde etmek için 50 veya daha geniş örnek büyüklüğü alınmasını önerir⁵¹.

Güvenirlik çalışması yapılması için gerekli örnek büyüklüğü tahminlerinin, literatürde yer alan, yukarıda da yer verilen örnek büyüklüğü önerileri yerine, araştırmacıdan alınacak planlanan Cronbach alfa değerine, ölçekte bulunan madde sayısına, istenilen güven aralığı genişliğine ve/veya istenilen güce bağlı olarak yapılması gerekir.



6. SONUÇ ve ÖNERİLER

Bu tez, ölçek geliştirme prosedürlerinde güvenirlik çalışmaları için gerekli örnek büyüklüğü miktarını araştırmaktadır. Buradaki temel amaç, “minimum maliyet ile maksimum bilgi kriteri felsefesi baz alınarak yapılan ölçek güvenirlik çalışmaları için örnek büyüklüğü ne olmalıdır?” sorusuna cevap aramaktır. Uygun örnek büyüklüğü önemlidir, zira bilindiği üzere küçük örnek büyüklüğü çalışmada düşük bir güce, büyük örnek büyüklüğü ise kaynak israfına yol açar.

Bu kapsamda, paralel ve paralel dışı ölçüm durumunda yapılan istenilen doğruluk için simülasyon çalışmasından elde edilen genel sonuçlar:

- Güven aralığı genişliği arttıkça örnek büyüklüğünün azalması,
- Planlanan Cronbach alfa düzeyi arttıkça örnek büyüklüğünün azalması,
- Madde sayısı arttıkça örnek büyüklüğünün azalmasıdır.

Bu sonuçlara ek olarak, madde sayısı 50 olduğunda bütün Cronbach alfa değer aralıkları ve güven aralığı genişlikleri için örnek büyüklüğü sabitlenmiştir.

İstenilen güç için yapılan simülasyon çalışmasından elde edilen genel sonuçlar:

- Planlanan Cronbach alfa düzeyi arttıkça örnek büyüklüğünün azalması,
- İstenilen güç düzeyi arttıkça örnek büyüklüğünün artması,
- Madde sayısı arttıkça örnek büyüklüğünün azalmasıdır.

Bu sonuçlara ek olarak, istenilen doğruluğa dayalı simülasyon çalışmasında olduğu gibi madde sayısı en geç 50’den sonra olduğunda spesifik güç değerleri ve belirlenmiş Cronbach alfa değer aralıkları için örnek büyüklükleri sabitlenmiştir. Takribi bir genelleme olarak, araştırmacılara yardımcı olması adına önerilen örnek büyüklüğü, yüksek madde sayısı varlığında ve çalışmalarda genellikle istenilen güç seviyesi olan %80 baz alınarak tablodaki gibi verilmiştir.

Planlanan Cronbach Alfa Değer Aralığı	Örnek Büyüklüğü(%80 Güç)
0.60-0.69	220
0.70-0.79	131
0.80-0.89	66
0.90-0.99	23

Bu tablonun bir benzeri %80 güce tekabül eden güven aralığı için de verilebilir, ki bu aralık takriben 0.18'dir. Zira, istenilen güç için örnek büyüklüğü ve istenilen doğruluk için örnek büyüklüğü çalışmaları harmanlanmak istenirse, istenilen doğruluk için örnek büyüklüğü çalışmasında ele alınan 0.15 güven aralığı genişliğine karşılık gelen örnek büyüklüklerinin, istenen güç için örnek büyüklüğü çalışmasında %85-%90 arası güç değerine karşılık gelen örnek büyüklüklerine denk geldiği görülmektedir. Buradaki kısıtlılık, çalışmalarda arzu edilen güç seviyesi olan %80'e ulaşmak adına istenilen doğruluk için örnek büyüklüğü çalışmasında gerekli güven aralığı genişliğine çalışmamızda yer verilmemesidir. Ancak, 0.15'ten daha geniş güven aralığı genişlikleri alınarak %80 güce ulaşmak mümkündür.

Bu tez çalışması kapsamında güvenilirlik çalışmalarında örnek büyüklüğü hesaplanmasında kullanılan iki yöntemle örnek büyüklüğü hesaplamalarının daha doğru yapılması ve araştırmacılara daha doğru bir yön vermek mümkün olabilir.

Ölçeklerde yapılan güvenilirlik hesaplamalarında, klasik test teorisi yerine modern test teorisi ve buna bağlı olarak Rasch analizine bağlı değerlendirmeler yapılması ve böyle bir senaryoda oluşturulacak olan örnek büyüklüğü tasarımları ilerleyen zamanlarda araştırma konusu olarak incelenebilir.

7. KAYNAKLAR

1. **Carmines EG, Zeller RA.** *Reliability and Validity Assessment.* 5th printing. Beverly Hills: Sage Publications Inc; **1982.**
2. **Şencan H.** *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik.* 1.Baskı, Ankara: Seçkin Yayıncılık, **2005.**
3. **Ercan İ, Kan İ** Ölçeklerde Güvenilirlik ve Geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, **2004**; 30: 211-216.
4. **Streiner DL, Norman GR, Cairney J.** *Health measurement scales: a practical guide to their development and use.* Oxford University Press, **2014.**
5. **Alpar R.** *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenilirlik.* Detay Yayıncılık, **2016.**
6. **Charter RA.** Study samples are too small to produce sufficiently precise reliability coefficients. *The Journal of General Psychology*, **2003**;130(2), 117-129.
7. **Bonett DG.** Sample size requirements for testing and estimating coefficient alfa. *Journal of educational and behavioral statistics*, **2002**; 27(4), 335-340.
8. **Bonett DG,Wright TA.** Cronbach's alfa reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, **2015**; 36(1), 3-15.
9. **Charter RA.** Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, **1999**; 21(4), 559-566.
10. **Iacobucci D, Duhachek, A.** Advancing alfa: Measuring reliability with confidence. *Journal of consumer psychology*, **2003**; 13(4), 478-487.
11. **Ponterotto JG, Ruckdeschel DE.** An overview of coefficient alfa and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and motor skills*, **2007**; 105(3), 997-1014.
12. **Terwee CB, Bot SD, de Boer MR , van der Windt DA , Knol DL , Dekker J, de Vet HC.** Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, **2007**; 60(1), 34-42.
13. **DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ , Lazzara DJ, Kostas-Polston E.** A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*, **2007**; 39(2), 155-164.
14. **McDowell I.** *Measuring health: a guide to rating scales and questionnaires.* Oxford university press, **2006.**
15. **De Gruijter DN, Leo JT.** *Statistical test theory for the behavioral sciences.* CRC Press. **2007.**
16. **Cohen RJ, Swerdlik ME, Phillips SM.** *Psychological testing and assessment: An introduction to tests and measurement.* Mayfield Publishing Co. **1996.**
17. **Erdoğan S.** Sürekli Değişkenler için Güvenilirlik Analizinde Kullanılan İlişki Katsayılarının Değişken Varyansından ve Etki Büyüklüğünden Etkilenme Durumları. Yüksek Lisans Tezi, Mersin Üniversitesi Sağlık Bilimleri Enstitüsü, Mersin, **2004.**

18. **Eisinga R, Grotenhuis MT, Pelzer B.** The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International journal of public health*, **2013**; 1-6.
19. **DeVellis RF.** *Scale development: Theory and applications* (Vol. 26). Sage publications. **2003**.
20. **Cronbach LJ.** Test “reliability”: Its meaning and determination. *Psychometrika*, **1947**; 12(1), 1-16.
21. **Vehkalahti K.** Reliability of measurement scales. *Statistical research reports*, **2000**; 17.
22. **Graham JM.** Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and psychological measurement*, **2006**; 66(6), 930-944.
23. **Drost EA.** Validity and reliability in social science research. *Education Research and Perspectives*, **2011**; 38(1), 105.
24. **Cook DA, Beckman TJ.** Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, **2006**; 119(2), 166-e7.
25. **Özdamar K.** *Eğitim, sağlık ve davranış bilimlerinde ölçek ve test geliştirme yapısal eşitlik modellemesi*. Nisan Yayıncılık, **2016**.
26. **Bannigan K, Watson R.** Reliability and validity in a nutshell. *Journal of clinical nursing*, **2009**; 18(23), 3237-3243.
27. **Lawshe CH.** A quantitative approach to content validity. *Personnel psychology*, **1975**; 28(4), 563-575.
28. **Lynn MR.** Determination and quantification of content validity. *Nursing research*, **1986**; 35(6), 382-386.
29. **Tavakol M, Dennick R.** Making sense of Cronbach's alfa. *International journal of medical education*, **2011**; 2, 53.
30. **Golafshani N.** Understanding reliability and validity in qualitative research. *The qualitative report*, **2003**; 8(4), 597-606.
31. **Schmitt N.** Uses and abuses of coefficient alfa. *Psychological assessment*, **1996**; 8(4), 350.
32. **Cortina JM.** What is coefficient alfa? An examination of theory and applications. *Journal of applied psychology*, **1993**; 78(1), 98.
33. **Streiner DL.** Starting at the beginning: an introduction to coefficient alfa and internal consistency. *Journal of personality assessment*, **2003**; 80(1), 99-103.
34. **Cronbach LJ.** Coefficient alfa and the internal structure of tests. *psychometrika*, **1951**; 16(3), 297-334.
35. **Yurdugül H.** Minimum Sample Size For Cronbach's Coefficient Alfa: A Monte-Carlo Study. *HU Journal of Education*, **2008**; 35, 397-405.
36. **Nunnally JD, Bernstein IH.** *Psychometric Theory*. New York, NY: McGraw Hill. **1994**.
37. **Van Zyl JM, Neudecker H, Nel DG.** On the distribution of the maximum likelihood estimator of Cronbach's alfa. *Psychometrika*, **2000**; 65(3), 271-280.

38. **Bonett DG.** Sample size requirements for comparing two alfa coefficients. *Applied Psychological Measurement*, **2003**; 27(1), 72-74.
39. **Cocchetti DV.** Sample size requirements for increasing the precision of reliability estimates: problems and proposed solutions. *Journal of Clinical and Experimental Neuropsychology*, **1999**; 21(4), 567-570.
40. **Köse İ.** Bir Üniversite Hastanesinde Çalışan Hemşirelerde Bel Ağrısının Fonksiyonel Yetersizlik Düzeyi ve Yaşam Kalitesine Etkisinin İncelenmesi Yüksek Lisans Tezi, Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü, Adana, **2017**.
41. **Fairbank JC, Pynsent PB.** The Oswestry disability index. *Spine*, **2000**; 25(22), 2940-2953.
42. **Ware Jr JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical care*, **1992**; 473-483.
43. **IBM Corp.** *IBM SPSS Statistics for Windows*, Version 23.0. Armonk, NY: IBM Corp., **2016**.
44. **R Studio Inc.** *R studio: Integrated Development Environment for R.*, Version 1.0.143, Boston, MA: R Studio Team, **2016**.
45. **Falk CF, Savalei V.** The relationship between unstandardized and standardized alfa, true reliability, and the underlying measurement model. *Journal of personality assessment*, **2011**;93(5), 445-453.
46. **Sijtsma K.** On the use, the misuse, and the very limited usefulness of Cronbach's alfa. *Psychometrika*, **2009a**; 74, 107–120.
47. **Green SB, Yang Y.** Commentary on coefficient alfa: A cautionarytale. *Psychometrika*, **2009**; 74, 121–135.
48. **Feldt LS.** The approximate sampling distribution of Kuder–Richardson reliability coefficient. *Psychometrika*, **1965**; 30,357–370.
49. **Feldt LS, Woodruff DJ, Salih FA.** Statistical inference for coefficient alfa. *Applied Psychological Measurement*, **1987**; 11, 93–103.
50. **Ercan I, Yazici B, Sigirli D, Ediz B, Kan I.** Examining Cronbach alfa, theta, omega reliability coefficients according to sample size. *Journal of Modern Applied Statistical Methods*, **2007**; 6(1), 27.
51. **Javali SB, Gudaganavar NV, Raj SM.** Effect of varying sample size in estimation of coefficients of internal consistency. *WebMedCentral BIostatistics*, **2011**; 2(2) WMC001572.

EKLER

R kodları

Simülasyon için Kodlar

#####İstenilen doğruluk için(Paralel ölçüm)

```
samplesize<-function(alfa,item,g,c1,c2){
  tilda<-seq(c1,c2,length.out = 1000)
  item<-item
  desiredwidth<-g
  alfa<-alfa
  satırsayısı<-length(tilda)*length(item)*length(desiredwidth)
  df<-matrix(0,nrow = satırsayısı,ncol=4)
  df[,2]<-rep(rep(desiredwidth,each=length(tilda)),length(item))
  df[,1]<-rep(item,each=length(desiredwidth)*length(tilda))
  df[,3]<-rep(tilda,length(item)*length(desiredwidth))
  i=0
  for( k in item){
    for(g in desiredwidth){
      for(pqtilda in tilda){
        #Cronbach alfa için güven aralığı
        #paralel ölçümler
        #örnek büyüklüğü
        #crntilda:planlanan cronbach alfa değeri
        # g:güven aralığı genişliği
        # k:madde sayısı
        zalfa <- qnorm(1 - alfa/2)
        no <- ceiling((8*k/(k - 1))*(1 - crntilda)^2*(zalfa/g)^2 + 2)
        c<- log(no/(no - 1))
        ul <- 1 - exp(log(1 - crntilda) - c + zalfa*sqrt(2*k/((k - 1)*(no - 2))))
        al <- 1 - exp(log(1 - crntilda) - c - zalfa*sqrt(2*k/((k - 1)*(no - 2))))
        wo<- ul - al
        ob <- ceiling((no - 2)*(wo/g)^2 + 2)
        i=i+1
        df[i,4]<-ob
      }
    }
  }
  return(df)
}
### width 0.05
#item5
dfitem_5_w5_0.60_0.69<-samplesize(0.05,5,0.05,0.60,0.69)
dfitem_5_w5_0.70_0.79<-samplesize(0.05,5,0.05,0.70,0.79)
dfitem_5_w5_0.80_0.89<-samplesize(0.05,5,0.05,0.80,0.89)
dfitem_5_w5_0.90_0.99<-samplesize(0.05,5,0.05,0.90,0.99)
#item 10
dfitem_10_w5_0.60_0.69<-samplesize(0.05,10,0.05,0.60,0.69)
dfitem_10_w5_0.70_0.79<-samplesize(0.05,10,0.05,0.70,0.79)
dfitem_10_w5_0.80_0.89<-samplesize(0.05,10,0.05,0.80,0.89)
dfitem_10_w5_0.90_0.99<-samplesize(0.05,10,0.05,0.90,0.99)
#item 20
dfitem_20_w5_0.60_0.69<-samplesize(0.05,20,0.05,0.60,0.69)
dfitem_20_w5_0.70_0.79<-samplesize(0.05,20,0.05,0.70,0.79)
dfitem_20_w5_0.80_0.89<-samplesize(0.05,20,0.05,0.80,0.89)
```

```

dfitem_20_w5_0.90_0.99<-samplesize(0.05,20,0.05,0.90,0.99)
#item 36
dfitem_36_w5_0.60_0.69<-samplesize(0.05,36,0.05,0.60,0.69)
dfitem_36_w5_0.70_0.79<-samplesize(0.05,36,0.05,0.70,0.79)
dfitem_36_w5_0.80_0.89<-samplesize(0.05,36,0.05,0.80,0.89)
dfitem_36_w5_0.90_0.99<-samplesize(0.05,36,0.05,0.90,0.99)
#item 50
dfitem_50_w5_0.60_0.69<-samplesize(0.05,50,0.05,0.60,0.69)
dfitem_50_w5_0.70_0.79<-samplesize(0.05,50,0.05,0.70,0.79)
dfitem_50_w5_0.80_0.89<-samplesize(0.05,50,0.05,0.80,0.89)
dfitem_50_w5_0.90_0.99<-samplesize(0.05,50,0.05,0.90,0.99)
#item 75
dfitem_75_w5_0.60_0.69<-samplesize(0.05,75,0.05,0.60,0.69)
dfitem_75_w5_0.70_0.79<-samplesize(0.05,75,0.05,0.70,0.79)
dfitem_75_w5_0.80_0.89<-samplesize(0.05,75,0.05,0.80,0.89)
dfitem_75_w5_0.90_0.99<-samplesize(0.05,75,0.05,0.90,0.99)
#item 100
dfitem_100_w5_0.60_0.69<-samplesize(0.05,100,0.05,0.60,0.69)
dfitem_100_w5_0.70_0.79<-samplesize(0.05,100,0.05,0.70,0.79)
dfitem_100_w5_0.80_0.89<-samplesize(0.05,100,0.05,0.80,0.89)
dfitem_100_w5_0.90_0.99<-samplesize(0.05,100,0.05,0.90,0.99)

##plot width 0.05

ornekbuyuklugu_w005<-
c(dfitem_5_w5_0.60_0.69[,4],dfitem_5_w5_0.70_0.79[,4],dfitem_5_w5_0.80_0.89[,4],dfitem_5_w5_0.90_0.99[,4],dfitem_10_w5_0.60_0.69[,4],dfitem_10_w5_0.70_0.79[,4],dfitem_10_w5_0.80_0.89[,4],dfitem_10_w5_0.90_0.99[,4],dfitem_20_w5_0.60_0.69[,4],dfitem_20_w5_0.70_0.79[,4],dfitem_20_w5_0.80_0.89[,4],dfitem_20_w5_0.90_0.99[,4],dfitem_36_w5_0.60_0.69[,4],dfitem_36_w5_0.70_0.79[,4],dfitem_36_w5_0.80_0.89[,4],dfitem_36_w5_0.90_0.99[,4],dfitem_50_w5_0.60_0.69[,4],dfitem_50_w5_0.70_0.79[,4],dfitem_50_w5_0.80_0.89[,4],dfitem_50_w5_0.90_0.99[,4],dfitem_75_w5_0.60_0.69[,4],dfitem_75_w5_0.70_0.79[,4],dfitem_75_w5_0.80_0.89[,4],dfitem_75_w5_0.90_0.99[,4],dfitem_100_w5_0.60_0.69[,4],dfitem_100_w5_0.70_0.79[,4],dfitem_100_w5_0.80_0.89[,4],dfitem_100_w5_0.90_0.99[,4])

maddesayısı<-
c(rep(5,4000),rep(10,4000),rep(20,4000),rep(36,4000),rep(50,4000),rep(75,4000),rep(100,4000))

Cronbachalfa<-
c(rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000))

graph<-data.frame(ornekbuyuklugu_w005,maddesayısı,Cronbachalfa)
library(Rmisc)
standarterror05 <- summarySE(graph, measurevar="ornekbuyuklugu_w005",
groupvars=c("maddesayısı","Cronbachalfa"))
graph$Cronbachalfa[graph$Cronbachalfa==1]<-"cron6069"
graph$Cronbachalfa[graph$Cronbachalfa==2]<-"cron7079"
graph$Cronbachalfa[graph$Cronbachalfa==3]<-"cron8089"
graph$Cronbachalfa[graph$Cronbachalfa==4]<-"cron9099"
plot05<-ggplot(standarterror05,aes(x=maddesayısı, y=ornekbuyuklugu_w005,colour
=as.factor(Cronbachalfa)))+
  geom_errorbar(aes(ymin=ornekbuyuklugu_w005-ci,ymax=ornekbuyuklugu_w005+ci),
width=3,size=0.5)+
  geom_line()+
  geom_point()+
  scale_x_continuous(breaks=c(5,10,20,36,50,75,100,200))+

```

```

scale_colour_hue(name="Cronbach alfa değeri aralığı",
  labels=c("0.60-0.69","0.70-0.79","0.80-0.89","0.90-0.99"))+
ylab("örnek büyüklüğü")+
xlab("madde sayısı")
plot05

```

#####İstenilen doğruluk için(Paralel dışı ölçüm)

```

rm(list = ls())
covmat<-function(cov,items,alfa,büyük,küçük){
  ###---Loadings
  fx<-rep(cov,items)
  ###---Error Variances
  err<-diag(runif(items,0.36,0.81))
  ###---matrix of factor covariances
  phi<-matrix(1, nrow=1)
  covmat<-(fx%*%phi%*%t(fx)+err)

  m <- matrix(rep(1, ncol(covmat)), nrow=ncol(covmat), ncol=1)
  c0 <- t(m)%*%covmat%*%m
  r1 <- sum(diag(covmat))
  r2 <- sum(diag(covmat%*%covmat))
  f1 <- c0^3
  f2 <- c0*(r2 + r1^2)
  f3 <- 2*r1*t(j)%*%(covmat%*%covmat)%*%j
  f4 <- 2*ncol(covmat)^2/(f1*(ncol(covmat) - 1)^2)
  crn <- (ncol(covmat)/(ncol(covmat) - 1))*(1 - r1/c0)
  if (crn < büyük && crn > küçük){
    return(list(covmat,c(crn,f2,f3,f4)))
  }
}
k=0
a<-list()
item<-c(5,10,20,36,50,75,100)

vector<-round(runif(100000,0.1,0.99),30)
syc=0
for(j in item){
  syc=syc+1
  for(i in vector){
    k=k+1
    b<-covmat(i,j,0.05,büyük=0.999,küçük=0.9)
    ##b'de eleman varsa a listesine k.eleman olarak b yi ekle
    if(is.null(b)==0){a[[k]]<-b} else {k=k-1}
    if(length(a)==syc*1000){break}
  }
}
CICron <- function(crn,g,VK,k,f2,f3,f4) {
  zalfa <- qnorm(1 - 0.05/2)
  no <- ceiling((8*ncol(VK)/(ncol(VK) - 1))*(1 - crn)^2*(zalfa/g)^2 + 2)
  c <- log(no/(no - 1))
  varyans <- f4*(f2 - f3)/(no-3)
  al <- 1 - exp(log(1 - crn) - c + zalfa*sqrt(varyans/(1 - crn)^2))
  ul <- 1 - exp(log(1 - crn) - c - zalfa*sqrt(varyans/(1 - crn)^2))
  wo <- ul-al
  ob<- ceiling((no - 2)*(wo/g)^2 + 2)
}

```

```

ciktı<- c(lowerlimit=al, cronbach=crn,
upperlimit=ul,width0=wo,samplesize=ob,width=g,item=ncol(VK))
return(ciktı)
}
width<-c(0.05,0.1,0.15)
sonuc<-matrix(0,length(a)*length(width),7)
k=0
for (j in width){
  for (i in 1:length(a)){
    k=k+1

sonuc[k,]<-CIcron(w=j,V=a[[i]][[1]],r=a[[i]][[2]][1],a2=a[[i]][[2]][2],a3=a[[i]][[2]][3],a4=a[[i]][[2]][4])
  }
}
sonuc<-as.data.frame(sonuc)
names(sonuc)<-c("lowerlimit","r","upperlimit","width0","samplesize","width","item")
head(sonuc)
library(Rmisc)
standarterror <- summarySE(sonuc, measurevar="samplesize", groupvars=c("item","width"))
plot4<-ggplot(standarterror,aes(x=item, y=samplesize,colour =as.factor(width) ))+
  geom_errorbar(aes(ymin=samplesize-ci,ymax=samplesize+ci), width=3,size=0.50)+
  geom_line( )+
  geom_point()+
  scale_x_continuous(breaks=c(5,10,20,36,50,75,100))+
  scale_colour_hue(name="Güven aralığı genişliği",
    labels=c("0.05","0.1","0.15"))+
  ggtitle("Planlanan Cronbach alfa değeri aralığı:0.90-0.99")+
  ylab("örnek büyüklüğü")+
  xlab("madde sayısı")
plot4<-plot4 + theme(
  plot.title = element_text( size=10, face="bold")
)

```

Gerçek Veri için Kodlar

#####ODI(doğruluk-paralel)

```

library(memisc)
ODI <-as.data.set(spss.system.file("C:/Users/hp/Desktop/oswestry disability index.sav"))
ODI<-as.data.frame(ODI)

```

```

library(psy)
cronODI<-cronbach(ODI)

```

####kovaryans

```

##20
k20<-replicate(1000,cov(sample(ODI,20, replace = TRUE, prob = NULL)))
o20<-apply(k20,1:2,mean)
s20<-apply(k20,1:2,sd)
LL20<-o20-(1.96*s20)
UL20<-o20+(1.96*s20)
##40
k40<-replicate(1000,cov(sample(ODI,40, replace = TRUE, prob = NULL)))
o40<-apply(k40,1:2,mean)
s40<-apply(k40,1:2,sd)
LL40<-o40-(1.96*s40)
UL40<-o40+(1.96*s40)

```

```

##80
k80<-replicate(1000,cov(sample(ODI,80, replace = TRUE, prob = NULL)))
o80<-apply(k80,1:2,mean)
s80<-apply(k80,1:2,sd)
LL80<-o80-(1.96*s80)
UL80<-o80+(1.96*s80)

##120
k120<-replicate(1000,cov(sample(ODI,120, replace = TRUE, prob = NULL)))
o120<-apply(k120,1:2,mean)
s120<-apply(k120,1:2,sd)
LL120<-o120-(1.96*s120)
UL120<-o120+(1.96*s120)

##160
k160<-replicate(1000,cov(sample(ODI,160, replace = TRUE, prob = NULL)))
o160<-apply(k120,1:2,mean)
s160<-apply(k120,1:2,sd)
LL160<-o160-(1.96*s160)
ULv<-o160+(1.96*s160)

##200
k200<-replicate(1000,cov(sample(ODI,200, replace = TRUE, prob = NULL)))
o200<-apply(k120,1:2,mean)
s200<-apply(k120,1:2,sd)
LL200<-o200-(1.96*s200)
UL200<-o200+(1.96*s200)

##güven aralığı

CIcon <- function(k,l,ob,VK) {

  zalfa <- qnorm(1 - 0.05/2)
  d <- log(ob/(ob - 1))
  m<- matrix(rep(1, l), nrow=l, ncol=1)
  ao<- t(m)%*%VK%*%m
  r1 <- sum(diag(VK))
  r2 <- sum(diag(VK%*%VK))
  p1 <- ao^3
  p2 <- ao*(r2 + r1^2)
  p3 <- 2*p1*t(m)%*%(VK%*%VK)%*%m
  p4 <- 2*k^2/(p1*(k - 1)^2)
  crn <- (k/(k - 1))*(1 - r1/ao)
  ul <- 1 - exp(log(1 - crn) - d - zalfa *sqrt(2*k/((k - 1)*(ob - 2))))
  al <- 1 - exp(log(1 - crn) - d + zalfa *sqrt(2*k/((k - 1)*(ob - 2))))
  wo <- ul - ll
  çıktı <- c(al, crn, ul,wo)
  return(çıktı)
}

##loop

set.seed(1912)
puren<-0
df<-matrix(0,nrow = 4,ncol=6)
nvector<-c(20,40,80,120,160,200)

for(i in nvector){
  puren<-puren+1

```

```

odi<-replicate(1000,CIcron(q=10,k=10,n=i,V=cov(sample(ODI, i, replace = TRUE, prob = NULL))))
df[,puren]<-apply(odi,1,mean,na.rm=T)
}

df
df<-data.frame(df)
names(df)<-c("n=20","n=40","n=80","n=120","n=160","n=200")
rownames(df)<-c("LL","r","UL","w0")
df

#### grafik
##cronbach

set.seed(1912)
puren<-0
cronODI<-matrix(0,nrow = 6,ncol=1000)
nvvector<-c(20,40,80,120,160,200)
for(i in nvvector){

  puren<-puren+1

  cronODI[puren,]<-replicate(1000,CIcron(10,10,i,cov(sample(ODI, i, replace = T, prob = NULL))))[2,]
}
cronODI<-data.frame(cronODI)
cronODI<-c(t(cronODI))

factor<-c(rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(5,1000),rep(6,1000))

veriODI<-cbind(cronODI,factor)
veriODI<-data.frame(veriODI)
veriODI$factor<-factor(veriODI$factor,label=c("n=20","n=40","n=80","n=120","n=160","n=200"))

####error bar ci'ye göre

örb<-c(rep(20,1000),rep(40,1000),rep(80,1000),rep(120,1000),rep(160,1000),rep(200,1000))
toplu<-cbind(cronODI,örb)
toplu<-data.frame(toplu)
toplu$örb<-factor(toplu$örb,label=c("n=20","n=40","n=80","n=120","n=160","n=200"))
library(ggplot2)
library(Rmisc)
standarterror <- summarySE(toplu, measurevar="cronODI", groupvars=c("örb"))

plot<-ggplot(standarterror,aes(x=örb, y=cronODI))+

  geom_errorbar(aes(ymin=cronODI-ci,ymax=cronODI+ci), width=0.3,size=0.30,colour="blue")+
  geom_line( )+
  geom_point(colour="blue")+

  scale_y_continuous(limits=c(0.8,0.91),breaks=c(0.80,0.85,0.90))+

  ylab("Cronbach alfa")+
  xlab("örnek büyüklüğü")
plot

```

```
#####SF-36(doğruluk-paralel dışı)

library(memisc)
SF36 <-as.data.set(spss.system.file("C:/Users/hp/Desktop/SF36puan.sav"))

SF36<-as.data.frame(SF36)
SF36<-SF36[1:36]

library(psy)
fizikselfonk<-cronbach(SF36[,3:12])
fizikselrolguclugu<-cronbach(SF36[,13:16])
emosyonelrolguclugu<-cronbach(SF36[,17:19])
enerji_ canlılık_ vitalite<-cronbach(SF36[,c(23,27,29,31)])
ruhsalsaglık<-cronbach(SF36[,c(24,25,26,28,30)])
sosyalislevsellik<-cronbach(SF36[,c(20,32)])
agrı<-cronbach(SF36[,c(21,22)])
genelsaglıkalgısı<-cronbach(SF36[,c(1,33,34,35,36)])

k20<-replicate(1000,cov(sample(SF36[,3:12],20, replace = TRUE, prob = NULL)))
o20<-apply(k20,1:2,mean)
s20<-apply(k20,1:2,sd)
LL20<-o20-(1.96*s20)
UL20<-o20+(1.96*s20)

##güven aralığı
C1cron <- function(alfa, k, ob, VK,l) {

  zalfa <- qnorm(1 - alfa/2)
  c <- log(ob/(ob - 1))
  m <- matrix(rep(1, l), nrow=1, ncol=1)
  ao <- t(m)%*%VK%*%m
  r1 <- sum(diag(VK))
  r2 <- sum(diag(VK%*%VK))
  p1 <- ao^3
  p2 <- ao*(r2 + r1^2)
  p3 <- 2*r1*t(m)%*%(VK%*%VK)%*%m
  p4 <- 2*k^2/(p1*(k - 1)^2)
  crn <- (k/(k - 1))*(1 - r1/ao)
  varyans <- p4*(p2 - p3)/(ob - 3)
  al <- 1 - exp(log(1 - crn) - c + zalfa *sqrt( varyans /(1 - crn)^2))
  ul <- 1 - exp(log(1 - crn) - c - zalfa *sqrt( varyans /(1 - crn)^2))
  dif<-ul-al
  çıktı <- c(al,crn, ul,dif)
  return(çıktı)
}

##loop
set.seed(1912)

puren<-0
df<-matrix(0,nrow = 4,ncol=5)
nvector<-c(10,20,40,80,120)
for(i in nvector){

  puren<-puren+1
```

```

FF<-replicate(1000,CIcron(0.05,36,i,cov(sample(SF36[,3:12]), i, replace = TRUE, prob = NULL)),36))
df[,puren]<-apply(FF,1,mean)
}

df
df<-data.frame(df)
names(df)<-c("n=10","n=20","n=40","n=80","n=120")
df

#### grafik
##cronbach

set.seed(1912)
puren<-0
cronFF<-matrix(0,nrow = 5,ncol=1000)
nvector<-c(10,20,40,80,120)
for(i in nvector){

  puren<-puren+1
  cronFF[puren,]<-replicate(1000,CIcron(0.05,36,i,cov(sample(SF36[,3:12]), i, replace = TRUE, prob =
NULL)),36))[2,]
}

cronFF<-data.frame(cronFF)
cronFF
cronFF<-c(t(cronFF))

factor<-c(rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(5,1000))

veri<-cbind(cronFF,factor)
veri<-data.frame(veri)
veri$factor<-factor(veri$factor,label=c("n=10","n=20","n=40","n=80","n=120"))

##errorbar
library(Rmisc)
veri<-cbind(cronFF,factor)
veri<-data.frame(veri)
veri$factor<-factor(veri$factor,label=c("n=10","n=20","n=40","n=80","n=120"))

özet<-summarySE(veri, measurevar="cronFF", groupvars="factor")
ffci<-özet$ci

ff<-ggplot(özet,aes(x=factor,y=cronFF))+

  geom_errorbar(aes(x=factor,ymin=cronFF-ci,ymax=cronFF+ci),width=0.2,size=1,colour="blue")+
  geom_point(colour="blue")+
  geom_line()+
  ylab("Güven aralığı genişliği")+
  xlab("örnek büyüklüğü")+
  scale_y_continuous(limits=c(0,0.1),breaks=c(0,0.05,0.1))
ff

ff+ggtitle("FF")

```

```

###ODI-güç

rm(list=ls())
library(memisc)
ODI <- as.data.set(spss.system.file("C:/Users/hp/Desktop/oswestry disability index.sav"))
ODI <- as.data.frame(ODI)

library(psy)
cronODI <- cronbach(ODI)

alfa <- 0.05
h <- 0.70

SSpow <- function(k,n,VK,l) {
  # alfa: alfa değeri
  # k: madde sayısı
  # crn: planlanan güvenirlik değeri
  # h: sıfır hipotezindeki güvenirlik değeri
  zalfa <- qnorm(1 - 0.05/2)
  m <- matrix(rep(1, l), nrow=1, ncol=1)
  ao <- t(m) %*% VK %*% m
  s1 <- sum(diag(VK))
  s2 <- sum(diag(VK %*% VK))
  p1 <- ao^3
  p2 <- ao*(s2 + s1^2)
  p3 <- 2*s1*t(m) %*% (VK %*% VK) %*% m
  p4 <- 2*k^2/(p1*(k - 1)^2)
  crn <- (k/(k - 1))*(1 - s1/ao)
  e <- (1 - crn)/(1 - h)

  c <- (n-2)*log(e)^2
  c <- sqrt(c)
  b <- 2*k/(k-1)
  b <- sqrt(b)
  za <- qnorm(1-0.05/2)
  zb <- (c/b)-za
  guc <- pnorm(zb)

  çıktı <- c(zb,crn,guc)
  return(çıktı)
}

##loop
set.seed(1912)
puren <- 0
pow <- matrix(0,nrow = 3,ncol=6)
nvector <- c(20,40,80,120,160,200)

for(i in nvector){

  puren <- puren+1
  powODI <- replicate(1000,sizePOWrel(10,i,cov(sample(ODI, i, replace = TRUE, prob = NULL)),10))
  pow[,puren] <- apply(powODI,1,mean,na.rm=T)
}

pow
pow <- data.frame(pow)
names(pow) <- c("n=20", "n=40", "n=80", "n=120", "n=160", "n=200")

```

```

#### grafik
##cronbach

set.seed(1912)
puren<-0
powODI<-matrix(0,nrow = 6,ncol=1000)
nvector<-c(20,40,80,120,160,200)
for(i in nvector){

  puren<-puren+1
  powODI[puren,]<-replicate(1000,sizePOWrel(10,i,cov(sample(ODI, i, replace = TRUE, prob =
NULL)),10))[3,]
}
powODI<-data.frame( powODI)

powODI<-c(t(powODI))
factor<-c(rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(5,1000),rep(6,1000))

veriODI<-cbind( powODI,factor)
veriODI<-data.frame(veriODI)
veriODI$factor<-factor(veriODI$factor,label=c("n=20","n=40","n=80","n=120","n=160","n=200"))

####error bar ci'ye göre
library(ggplot2)
örb<-c(rep(20,1000),rep(40,1000),rep(80,1000),rep(120,1000),rep(160,1000),rep(200,1000))
toplu<-cbind( powODI,örb)
toplu<-data.frame(toplu)

toplu$örb<-factor(toplu$örb,label=c("n=20","n=40","n=80","n=120","n=160","n=200"))

library(Rmisc)
standarterror <- summarySE(toplu, measurevar="powODI", groupvars=c("örb"))
plot<-ggplot(standarterror,aes(x=örb, y= powODI))+

  geom_errorbar(aes(ymin= powODI-ci,ymax= powODI+ci), width=0.3,size=0.30,colour="blue")+
  geom_line( )+
  geom_point(colour="blue")+
  scale_y_continuous(limits=c(0.40,1),breaks=c(0.40,0.60,0.80,0.90,0.95,0.99))+
  ylab("Güç")+
  xlab("örnek büyüklüğü")
plot

####SF-36-güç

library(memisc)
SF36 <-as.data.set(spss.system.file("C:/Users/hp/Desktop/SF36puan.sav"))
SF36<-as.data.frame(SF36)
SF36<-SF36[1:36]
library(psy)
cronbach(SF36)
alfa<-0.05
h<-0.80

```

```

SSpow<- function(k,n,VK,l) {
  # alfa: alfa değeri
  # k: madde sayısı
  # crn: planlanan güvenirlik değeri
  # h: sıfır hipotezindeki güvenirlik değeri
  zalfa <- qnorm(1 - 0.05/2)
  m <- matrix(rep(1, l), nrow=1, ncol=1)
  ao <- t(m)%*%VK%*%m
  s1 <- sum(diag(VK))
  s2 <- sum(diag(VK%*%VK))
  p1 <- ao^3
  p2 <- ao*(s2 + s1^2)
  p3 <- 2*s1*t(m)%*%(VK%*%VK)%*%m
  p4 <- 2*k^2/(p1*(k - 1)^2)
  crn <- (k/(k - 1))*(1 - s1/a0)
  e <- (1 - crn)/(1 - h)

  c<-(n-2)*log(e)^2
  c<-sqrt(c)
  b<-2*(1/(1-1))
  b<-sqrt(b)
  za<-qnorm(1-0.05/2)
  zb<-(c/b)-za
  guc<-pnorm(zb)

  çıktı<-c(crn,guc)
  return(çıktı)
}

##loop
set.seed(12)
puren<-0
pow<-matrix(0,nrow = 2,ncol=5)
nvector<-c(10,20,40,80,120)

for(i in nvector){

  puren<-puren+1
  powSF<-replicate(1000,sizePOWrel(36,i,cov(sample(SF36, i, replace = TRUE, prob = NULL))))
  pow[,puren]<-apply(powSF,1,mean,na.rm=T)

}

pow
pow<-data.frame(pow)
names(pow)<-c("n=10","n=20","n=40","n=80","n=120")

set.seed(155555)
puren<-0
powFF<-matrix(0,nrow = 5,ncol=1000)
nvector<-c(10,20,40,80,120)

for(i in nvector){

  puren<-puren+1
  powFF[puren,]<-replicate(1000,sizePOWrel(36,i,cov(sample(SF36, i, replace = TRUE, prob =
NULL)),36))[2,]
}

```

```

powFF<-data.frame(powFF)
powFF<-c(t(powFF))

factor<-c(rep(1,1000),rep(2,1000),rep(3,1000),rep(4,1000),rep(5,1000))

veriFF<-cbind(powFF,factor)
veriFF<-data.frame(veriFF)
veriFF$factor<-factor(veriFF$factor,label=c("n=10","n=20","n=40","n=80","n=120"))

##errorbar
library(ggplot2)
örb<-c(rep(10,1000),rep(20,1000),rep(40,1000),rep(80,1000),rep(120,1000))
toplu<-cbind(powFF,örb)

toplu<-data.frame(toplu)
toplu$örb<-factor(toplu$örb,label=c("n=10","n=20","n=40","n=80","n=120"))

library(Rmisc)
standarterror <- summarySE(toplu, measurevar="powFF", groupvars=c("örb"))

plot1<-ggplot(standarterror,aes(x=örb, y= powFF))+
  geom_errorbar(aes(ymin=powFF-ci,ymax=powFF+ci), width=0.3,size=0.30,colour="blue")+
  geom_point(colour="blue")+
  scale_y_continuous(limits=c(0.80,1.01),breaks=c(0.80,0.90,0.99))+
  ylab("Güç")+
  xlab("örnek büyüklüğü")
plot1
plot1<-plot1+ggtitle("SF36")

```

ÖZGEÇMİŞ

1991 yılında Adana’da doğdu. Üniversiteye kadar olan eğitimini Adana’da tamamladıktan sonra 2009 yılında girdiği Ege Üniversitesi İstatistik Bölümü’nden 2014 yılında mezun oldu. 2015 yılı güz döneminde, Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü’ne bağlı Biyoistatistik Anabilim Dalı’nda Yüksek Lisans öğrenimine başladı. Aynı yıl, Çukurova Üniversitesi Fen Bilimleri Enstitüsü’ne bağlı İstatistik Bölümü’nde de Yüksek Lisans öğrenimine başladı. 2017 yılında Çukurova Üniversitesi Biyoistatistik Anabilim Dalı’na Araştırma Görevlisi olarak atandı ve dört aydır görevine devam etmektedir.