

SOSYAL BİLİMLERDE İSTATİSTİK II



PROF.DR.YÜKSEL TERZİ

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN-EDEBİYAT FAKÜLTESİ

İSTATİSTİK BÖLÜMÜ

SAMSUN

2021

4. GÜVENİLİRLİK

Bir ölçme aracında (testte veya ankette) bütün soruların birbirleriyle tutarlılığını, ele alınan oluşumu ölçmede türdeşliğini ortaya koyan bir kavramdır.



- Ölçme araçlarının güvenilirliğini değerlendirmek amacıyla geliştirilmiş yöntemlere Güvenirlik Analizi ve bu araçta yer alan soruların irdelenmesine ise soru analizi (Reliability and Item Analysis) denilmektedir.



- Testlerin gvenirliđini analiz etmek amacıyla gvenirlik katsayıları hesaplanmaktadır. Bugne kadar gvenirlik analizi iin bir ok yntem geliřtirilmiřtir.



Güvenilirlik (reliability) ve geçerlilik (validity) analizleri yapılmadan bir arařtırmanın analiz sonuçlarını yorumlamak doęru deęildir. Güvenilirlik ve geçerlilik herhangi bir řeyin uygun ve saęlam olduęu hakkında bize bilgi verir. Örneęin ayakkabının ayaęımıza hem uygun olması hem ayakkabının saęlam olması gerekir. Burada uygunluk geçerlilięi, saęlamlık ise güvenilirlięi temsil etmektedir.

Geçerlilik, bir testin ölçmek istediği özelliği ölçmesi ile ilgilidir. Bir test ölçmek istediği özelliği doğru ve diğer özelliklerle karıştırmadan ölçüyor ise bu testin geçerli olduğu söylenir. Geçerli bir testte bulunması gereken en önemli özellik testin güvenilir olmasıdır.

Güvenilirlik bir ölçüm sürecinde ölçüm işleminin tekrarlanabilirliği ya da tekrarlardaki tutarlılıktır.

- Güvenilirlik bir özelliğin bağımsız ölçümleri arasındaki kararlılıktır.**
- Güvenilirlik zamana göre değişmezlik ölçüsüdür.**

Güvenilirlik bir testin geçerliliğini etkiler. Geçerli bir test muhakkak güvenilir olmalıdır. **Güvenilir bir test geçerli olmayabilir. Yani güvenilirlik geçerliliği garanti etmemektedir.**

Örneğin kol kuvvetini ölçmek için kullanılan barfix testinin güvenilirliği çok yüksektir. Bu test yanlışlıkla bacak kuvvetini ölçmede kullanılırsa geçerli olmayacaktır. Yine 20-45 yaş arasındaki bayanlar için geliştirilen ve belli bir özelliği ölçen güvenilirliği yüksek bir ölçeğin 55 yaş üzerinelere uygulanırsa geçerli olmayacaktır.

Araştırma çalışmaları yapılırken nesnelerin değişik özellikleri ile ilgili ölçüm yapmak için değişik aletler, laboratuvar aygıtları kullanılır veya sosyal bilimler, davranış bilimleri gibi bilim dallarında kullanılan anket çalışmaları veya psikolojik testlerde de farklı ölçme aygıtları kullanılır.

Bir araştırma sonucu elde edilen verilere sadece verinin iyi toplandığı için güvenilmez, yöntem kısmında araştırmada kullanılan aletlerin detay tanımları verilir, kullanım amaçları açıklanır, soru biçimleri ve ölçme ölçeği (adlandırma sıralama, aralık veya oran) açıklanır. Bunların hepsi geçerlilik ve güvenirliliğin sağlanabilmesi amacına yönelik gayretlerdir.

Bir arařtırmada geerlilik ve gvenirlilięi artırmak iin ařaęıdaki dzenlemelerin yapılmasına dikkat edilir (Thomas RK, Mueller RO.)

Sıra no	Yapılması gerekenler	Hangi Blmde Yapılır
1	alıřmanın hedef ve hipotezleri yalın ve sade bir Őekilde tanımlanmalıdır.	Giriř, Yntem
2	alıřmada kullanılan her aygıt yeterli detayda tanımlanmalıdır. Kullanılan aygıtların alıřmanın amalarına uygunluęu aıklanmalıdır. Uygun geerlilik delilleri toplanmalıdır. alıřmanın amacına uygun mevcut toplanan veri setinin destekledięi geerlilięin eřidi belirlenmelidir.	Yntem ve sonular
3	Gvenirlik ve geerlilięi artırıcı alınan nlemler aıklanmalıdır. nceki alıřmalardaki gvenirlik indeks deęerleri ile bu alıřmanın amalarına uygun Őekilde hesaplanan indeksler deęerler karřılařtırılmalıdır.	Yntem ve sonular
4	Arařtırmada uygulanan gvenirlik ve geerlilik alıřmaları yorumlanmalı ve rapor edilmelidir. alıřmanın sonuları bu mevcut deliller kapsamında tartiřılmalıdır.	Tartıřma

Geçerlilik ve güvenilirlik ölçme aygıtlarının önemli iki özelliğidir.

Güvenirlik ölçümün kararlılığı ile ilgili bir özelliktir. Alınan ölçümlerin kararlı olması veya değişmemesi istenir, bu kararlılık zamandan zamana, mekandan mekana, sorudan soruya, hakemden hakeme değişmeyen bir özelliktir.

Geçerlilik ise ölçmek istenen özelliğin, ölçülmesi niyet edilen hedef değerinden gerçekten ölçülebilmesi ile ilgili bir özelliktir. Dolayısıyla geçerlilik çalışmanın amacı ve araştırma sorusu (hipotez) ile yakından ilişkili bir husustur.

Araştırmacılar ölçme aracının farklı kültürler, farklı durumlar, farklı koşullarda aynı hassasiyetle ölçüm yapıp yapmadığı ile yakından ilgilidir. Örneğin Asya ırkından çocuklarda oldukça yüksek güvenilirlik ve geçerliliğe sahip bir psikolojik test, Afrika ırkından çocuklarda olmayabilir.

Güvenirlik ve geçerlilik kavramları ve tanımlamaları zaman içerisinde sürekli gelişmekte bunları ifade edebilmek için farklı ölçütler bulunmaktadır.

Geçerlilik mi, Güvenirlilik mi?

Geçerlilik mi daha önemli güvenirlilik mi daha önemli sorusunun yanıtını vermek biraz zordur.

Geçerli olmayan bir test ne kadar güvenilir olursa olsun bir işe yaramaz. Geçerlilik ön plana çıkmaktadır. En azından yapı ve içerik geçerliliği mutlaka sağlanmalıdır. Bir bakıma güvenirlilikte geçerliliğin bir parçası şeklinde düşünmek kavramların ne kadar iç içe olduğunu vurgulamak için yerinde olur. Burada ayrı ayrı konu edilmesinin amacı aralarındaki farklılıkları vurgulamaktır.

Geçerlilik ve Güvenirlik İlişkisi

Güvenirlik, geçerlik için bir ön koşuldur. Bir ölçme aracının güvenilir olması her zaman geçerli olduğu anlamına gelmez. Geçerlilik, güvenilirliği kapsayan bir kavramdır.

- Geçerlilik ile güvenilirlik arasındaki gerçek fark bir tanım meselesidir.
- **Güvenirlik yapılan ölçümün kararlılığını (consistency) veya daha basit olarak aynı şartlarda, aynı nesnelere, değişik zamanlarda veya kişilerce ölçüldüğünde benzer sonuçların bulunması hadisesidir.**
- **Geçerlilik ise ölçmek istenen, amaçlanan şeyin gerçekten ölçülüp ölçülmediği ile ilgilidir yani kesinlik (accuracy) ölçüsüdür.**

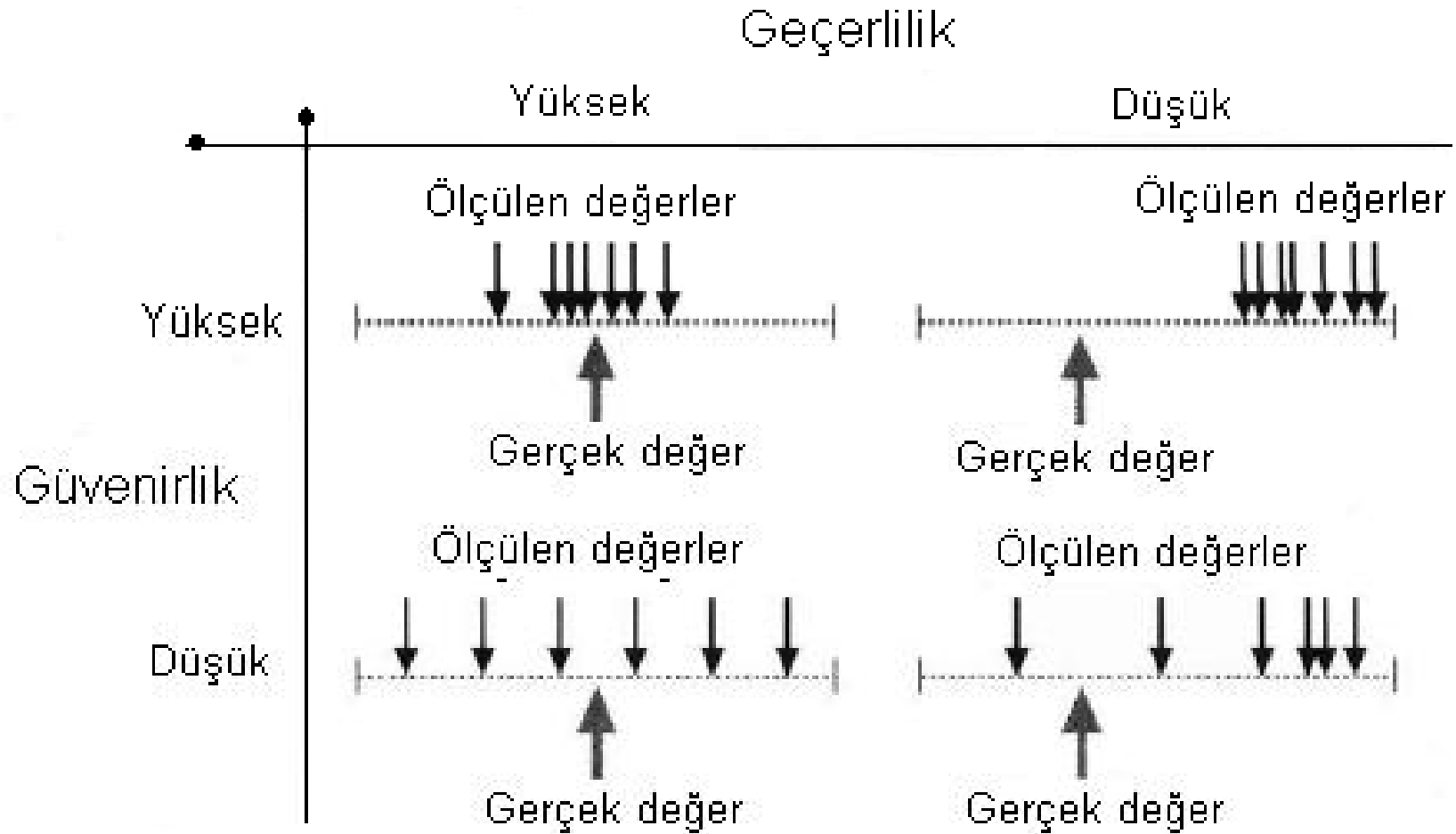
- Amaçlanan hedef için ölçüm yapılmıyorsa ne kadar hassas ölçüm yapılması fazlada bir önem taşımaz, o nedenle geçerlilik daha ön planda düşünülmesi gereken konudur.
- Geçerlilik her türlü hatadan etkilenmesine rağmen, güvenilirlik yalnızca tesadüfî hatadan etkilenir. Bir testin geçerliliği güvenilirlikten etkilenir, bu yüzden testin geçerlilik katsayısı, güvenilirlik katsayısının karekökünden daha büyük olamaz.

Bir testte hata kaynakları kontrol altına alındığı zaman testin güvenilirliği artar. Güvenirliliği artan bir testin geçerliliği de yükselir, ancak bu iki kavram birbirlerinden farklı kavramlardır.

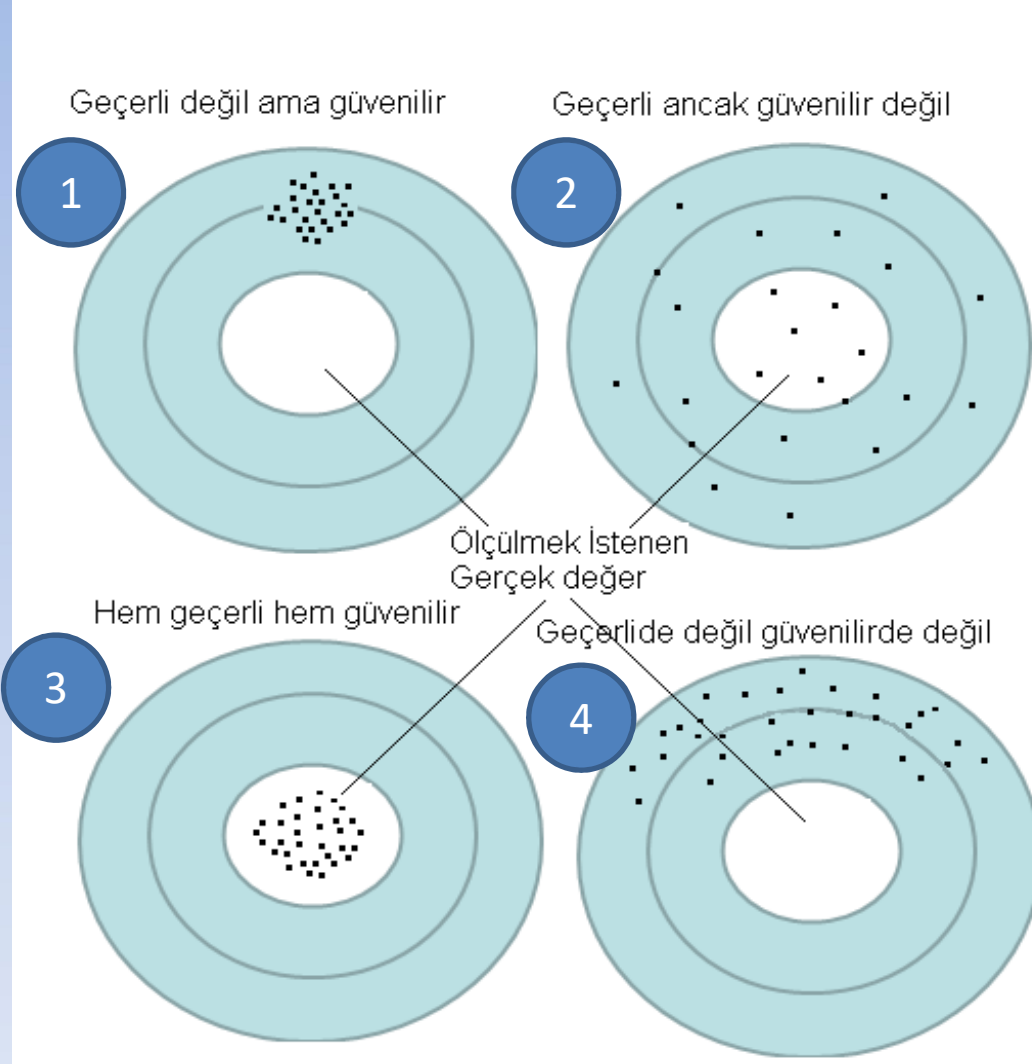
Güvenirlik bir özellik birkaç defa aynı araçla ölçüldüğünde her seferinde benzer sonuçlar elde edilmesi ile ilişkili bir kavramdır.

Geçerlilik ve Güvenirlik İlişki Şeması

Geçerlilik ve Güvenirlik



Kavramların şekille ifadesi



1. Her zaman aynı sonucu alıyor, ama istenen sonucu tam kapsamıyor.
2. İstenen sonucun tümünü kapsıyor, ama her zaman farklı sonuç alabiliyor. (Tesadüfî hata yüksek)
3. Her zaman aynı sonucu alıyor ve isteneni tam ve kesin kapsıyor. (Tesadüfî hata düşük)
4. Hem istenen sonucu tam kapsamıyor, hem de her zaman aynı sonucu alamıyor.

GÜVENİLİRLİK (Reliability)

Güvenilirlik, bir testin veya ölçeğin ölçmek istediği şeyi tutarlı ve istikrarlı bir biçimde ölçme derecesidir. Diğer bir deyişle, bireylerin test maddelerine verdikleri cevaplar arasındaki tutarlılıktır.

Güvenilir bir test veya ölçek, benzer şartlarda tekrar uygulandığında aynı benzer sonuçlar verir. Bir test veya ölçek ne derece güvenilir ise ondan elde edilen veriler de o derece güvenilirdir. Güvenilir olmayan bir ölçek yardımıyla elde edilen veriler faydasızdır. Mesela bir zeka testinde bir öğrenci bir gün 100, diğer gün 140 puan alıyorsa yapılan bu testin güvenilirliğinden bahsedilemez.

Güvenirlik;

- **Aynı şeyin bağımsız ölçümleri arasındaki kararlılıktır;**
- **Ölçülmek istenen belli bir şeyin, sürekli olarak aynı sembolleri almasıdır;**
- **Aynı süreçlerin izlenmesi ve aynı ölçütlerin kullanılması ile aynı sonuçların alınmasıdır;**
- **Ölçmenin, tesadüfî yanılılardan arınık olmasıdır;**
- **Bir ölçüm sürecinde ölçüm işleminin tekrarlanabilirliği, kararlılığı ya da tekrarlardaki tutarlılıktır.**

GÜVENİLİRLİK

- ✓ Endüstri mühendisine göre güvenilirlik: "üretimde hata oranının veya başarısızlık oranının düşük çıkmasıdır".
- ✓ Bir sosyologa göre güvenilirlik: "ölçüm sonuçlarının farklı anakütlelerde veya aynı ana kütleyle ait farklı örnek kütlelerde benzer sonuçlar vermesidir".
- ✓ Bir insan kaynakları yöneticisine göre güvenilirlik: "psikometrik test sonuçlarının uygulandığı farklı zaman dilimlerinde benzer sonuçlar vermesi ve iş yaşamında personelin göstereceği performansı doğru tahmin etmesidir".

Pilot araştırma verilerinde yapılan güvenilirlik analizleri bir ön yordama niteliğindedir. Güvenilirlik analizleri ise esas araştırma sonuçlarına göre yapılmalıdır.

Güvenilirliği sorgulama :

- **Bir ölçüm aracındaki maddelerin aynı kavramsal yapıyı hatasız bir biçimde ölçmesi**
- **Farklı zamanlarda yapılan ölçüm sonuçlarının benzer çıkması**
- **Bir ölçüm aracına ait sonuçların aynı kavramsal yapıyı ölçen diğer ölçüm araçlarının sonuçlarıyla tutarlı olması**
- **Farklı gözlemciler tarafından yapılan ölçüm sonuçlarının benzer çıkması**

Örnek

Güvenirlilik: (üç senaryo düşünelim)

1. Ağaç masamın boyutları bir zamandan diğerine değişmeyeceğine göre veya bir yerden diğer yere göre değişmeyeceği için ve benim çelik ölçme aletimin boyu değişmeyeceği için ölçüm nerede veya ne zaman yapılırsa yapılırsın fark etmez aynı bulunması gerekir. Eğer masa bir yerde 120 cm ölçülmüş ise başka bir yerde de veya başka bir günkü ölçümde de 120 cm gelecektir.

Birbiri peşine 3 ölçüm yapsak her 3 ölçümde de 120 cm bulunacaktır.

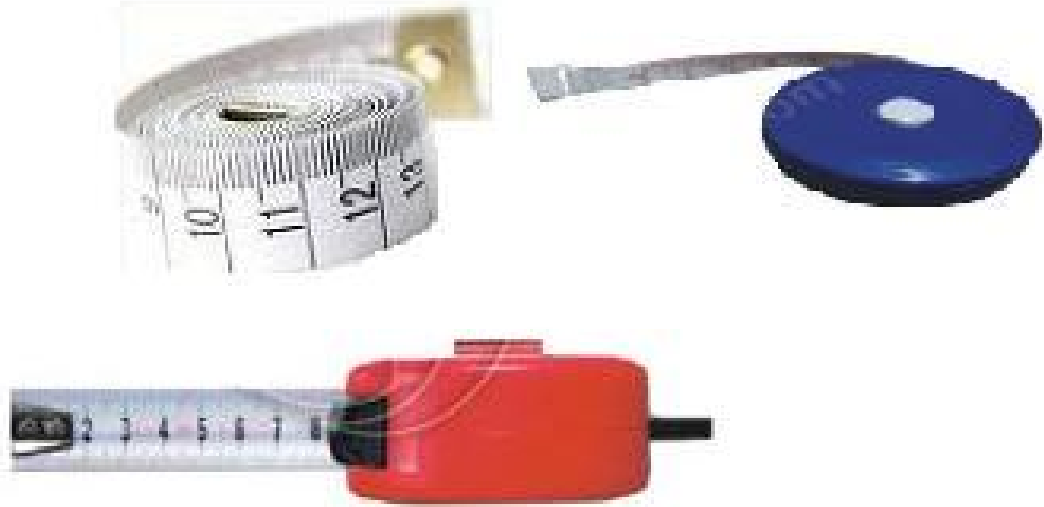
2. Diğer bir senaryoya göre ölçüm aletinin esneyebilen bezden yapıldığını ve çekme ile biraz uzayabildiğini varsayalım bu durumda 3 farklı ölçümde değişik ölçümler bulunacaktır, 120 cm, 121 cm ve 120,7 cm gibi.

3. Bir başka senaryoya göre ölçüm aleti sıcaklıktan daha fazla etkilenen daha esnek bir materyalden (plastik) yapılmış olsun. Bu aletle aynı masa oda ısısının farklı olduğu zamanlarda alındığında alet sıcaktan çok etkileneceği için çok daha farklı sonuçlar elde edilecektir. 120 cm, 150 cm ve 80 cm bulunmuş olsun.

Bu kavram ölçümün güvenilirliği ile ilgilidir. **Aynı nesne, aynı işlemle ölçüldüğünde neticenin aynı olması gerekir.** 1. durumda olduğu gibi alınan bir ölçüm **güvenilir** bir ölçümdür. 2. durumda çok az değişiklik vardır, dolayısıyla bu ölçümde **nispeten güvenilir**dir, ama 3. durumda çok büyük farklılıklar olduğu için ölçümler **güvenilirlikten uzaktır.**

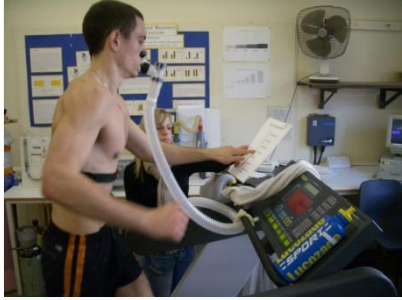




Nesne: Masa



Ölçüm aracı: Çelik metre
Mezura (Kot)
Mezura (plastik)

ÖRNEK1. 1. ve 2. arařtıřıcılar, altın standartla (gerçek deęere en yakın ölçüm) benzer ölçümler yapmıřtır. **Deęerlendiriciler arası güvenilirlik** tamdır.

	Altın standart Ölçüm	1. Arařtırıcı	2. Arařtırıcı
			
1. Kiři	60 ml.kg ⁻¹ .min ⁻¹	60 ml.kg ⁻¹ .min ⁻¹	60 ml.kg ⁻¹ .min ⁻¹
2. Kiři	55 ml.kg ⁻¹ .min ⁻¹	55 ml.kg ⁻¹ .min ⁻¹	55 ml.kg ⁻¹ .min ⁻¹
3. Kiři	70 ml.kg ⁻¹ .min ⁻¹	70 ml.kg ⁻¹ .min ⁻¹	70 ml.kg ⁻¹ .min ⁻¹

ÖRNEK2. 1. ve 2. arařtıřıcılar, altın standartla (gerçek deęere en yakın ölçüm) benzer ölçümler yapmıřtır. **Deęerlendiriciler arası güvenilirlik** tamdır. Ancak ölçümler hatalıdır.

	Altın standart Ölçüm	1. Arařtırıcı	2. Arařtırıcı
			
1. Kiři	60 ml.kg ⁻¹ .min ⁻¹	65 ml.kg ⁻¹ .min ⁻¹	65 ml.kg ⁻¹ .min ⁻¹
2. Kiři	55 ml.kg ⁻¹ .min ⁻¹	60 ml.kg ⁻¹ .min ⁻¹	60 ml.kg ⁻¹ .min ⁻¹
3. Kiři	70 ml.kg ⁻¹ .min ⁻¹	75 ml.kg ⁻¹ .min ⁻¹	75 ml.kg ⁻¹ .min ⁻¹

ÖRNEK3. 1. ve 2. arařtıřıcılar, altın standartla (gerçek deęere en yakın ölçüm) benzer ölçümler yapmıřtır. **Deęerlendiriciler arası güvenilirlik** yoktur, hem de ölçümler hatalıdır.

	Altın standart Ölçüm	1. Arařtırıcı	2. Arařtırıcı
			
1. Kiři	60 ml.kg ⁻¹ .min ⁻¹	72 ml.kg ⁻¹ .min ⁻¹	57 ml.kg ⁻¹ .min ⁻¹
2. Kiři	55 ml.kg ⁻¹ .min ⁻¹	61 ml.kg ⁻¹ .min ⁻¹	52 ml.kg ⁻¹ .min ⁻¹
3. Kiři	70 ml.kg ⁻¹ .min ⁻¹	40 ml.kg ⁻¹ .min ⁻¹	84 ml.kg ⁻¹ .min ⁻¹

GÜVENİLİRLİK ANALİZİ YÖNTEMLERİ

- 1. Test-Yeniden Test (Test-Retest)**
- 2. Paralel Formlar Yöntemi**
- 3. Gözlemciler Arası Güvenilirlik**
- 4. İç Tutarlılık Güvenilirliği**

GÜVENİLİRLİK ANALİZİ YÖNTEMLERİ

1. Test-Yeniden Test (Test-Retest Reliability Analysis)

Buna göre anket formu iki farklı zamanda (2-4 hafta veya 10-20 gün) aynı denekler üzerinde aynı şartlar altında uygulanır. İki farklı zamandaki gözlemler arasındaki korelasyon, güvenilirliğin bir göstergesi olacaktır. Bu ölçeğin bazı sorunları vardır. Önceki sonuçların sonraki sonuçları etkiyebilmesi, tekrar ölçme zorluğu gibi sıkıntılar olabilir.

Bir testin belli bir süre arayla aynı kişilere tekrar uygulanır. Toplam skor puanları ve her bir sorunun tutarlılığı Pearson-Spearman korelasyon ile değerlendirilir. Uyumsuz cevap alınan sorular anketten çıkarılır. Bunun için korelasyon değeri $+0,3$ 'ün altında olan sorular uyumsuz soru olarak kabul. Her bir alt ölçek için test re-test için korelasyonlara bakılır. Bulunan korelasyon pozitif ve 1'e yakınsa testin güvenilir olduğuna karar verilir.

2. Paralel Formlar Yöntemi

Bu yöntemle güvenilirlik tahmini için, bir testin en az iki eşdeğer formunun geliştirilmiş olması gerekir. Paralel iki testin eşdeğer olması için, her iki test, içindeki madde sayısı, niteliği ve ölçtükleri davranış bakımından birbirine denk olmalıdır.

Bu yöntem aynı gruba uygulanır ve iki ayrı formdan almış oldukları puanlar arasındaki korelasyona bakılır. Her iki test peş peşe uygulanabileceği gibi belli bir ara ile de uygulanabilir.

Bu yöntemle elde edilen güvenilirlik katsayısı, diğer yöntemlerle elde edilenlerden daha küçük çıkar.

3. Gzlemciler Arası Gvenilirlik

Gzlemciler nceden belirlenmiř bir puanlama sistemine baęlı olarak belirli bir olguyu, baęımsız bir řekilde deęerlendirebilirler. Bu deęerlendirmede verilen puanların birbirine benzer olması sz konusu puanların gvenilir olduęunu gsterir.

Birden ok gzlemcinin, birbirinden baęımsız olarak, aynı řeyleri lmeye alıřtıkları durumlarda uygulanan bir gvenilirlik ltdr. Baęımsız gzlemciler arası uyumu hesaplamak iin Korelasyon - Kendall's Coefficient of Concordance kullanılır.

lm aracı kullanılarak yapılan deęerlendirmede gzlemciler arası uyuřmanın en az 0,70 olması gerekir.

Gözlemcilerin verdikleri puanlar arasındaki uyumu belirlemek için verilerin niteliği göz önünde bulundurulur.

Gözlemci Puanlarında Güvenilirlik Analizleri

Ölçme Düzeyi	İki Gözlemci	İkiden Çok Gözlemci
(Kesikli Değişkenler)		
Nominal	- Uyum İndeksi - Cohen Kappa - Phi Katsayısı	- Uyum İndeksi - Cohen Kappa
Ordinal	-Spearman Sıra Kor. -Kendall Tau (a,b,c)	-Kendall W Uyum Katsayısı
Sürekli Değişken		
Eşit aralıklı /Oran	Pearson Korelasyon An. (Normal Dağılımlı)	Küme İçi Korelasyon Analizi (Intraclass Correlation Coefficient) (Normal Dağılımlı)

4. İ Tutarlılık (Internal consistency)

Bu yaklaşımda kavramı ölçerken çok sayıda maddeden oluşan (Likert tipi) bir ölçek kullanılarak, ölçekte yer alan maddeler arasındaki korelasyona bakarak içsel tutarlılık analizi yapılır. İ tutarlılık yönteminin amacı bir testin her bir sorusunun aynı niteliği ne kadar ölçtüğünü hesaplamaktır. İ tutarlılık Cronbach alpha, split half, Guttman, Paralel ve Kesin Paralel yöntemler ile hesaplanır. Sonuç 1'e ne kadar yakınsa iç tutarlılık o kadar yüksektir.

Ölçeğin tümü ve alanları için korelasyona dayalı madde analizi yapılır. Hotelling's T^2 analizi ($p < 0,05$ ise ortalama olarak ölçek maddeleri arasında fark var demektir.) ve yarı-test (split half, guttman) güvenilirlik sonuçları elde edilir ve Cronbach Alpha katsayıları hesaplanır. Her bir soru ile total skor puanları arasında korelasyon değeri $+0,3$ 'ün altında olan sorular ölçekten çıkarılır.

Ölçeğin tamamına ait iç tutarlılığa Cronbach Alpha katsayısı ile bakılır. Bazı sorular negatif, bazıları pozitif olur ve bunların Cronbach Alpha güvenilirlik katsayısına ayrı ayrı bakılır.

Güvenilirlik katsayısının hesaplanmasında Spearman ve Guttman değerleri hesaplanır. Bu değerlere “iç tutarlık katsayıları” denir. Güvenilirlik analizi için bu yöntemlerin kullanımında ölçülecek konuya ilişkin ölçekteki soru sayısının $k > 20$ ve denek sayısının $n > 50$ 'den fazla olmasına dikkat edilir. Güvenilirlik katsayılarının 0.60'ın üstünde olması gerekmektedir (Ural ve Kılıç,2005). Bir testin iç tutarlılık katsayısının 0,90'nın üstünde olması, o testin mükemmel güvenilirlikte olduğunu göstermektedir.

İçsel tutarlığı ölçmede çeşitli yaklaşımlar vardır.

a) Yarıya Bölme Yöntemi (Split Half)

Bu yöntem, tek bir test formu, tek bir öğrenci grubu ve tek bir test uygulaması gerektirir. Bu yöntemle test güvenilirliğini tahmin etmede, uygulanmış olan test iki eşdeğer yarıya bölünerek öğrencilerin testin iki yarısından aldıkları puanlar arasındaki korelasyon hesaplanır ve daha sonra bu hesaplanan korelasyondan hareketle Spearman-Brown formülünden de yararlanarak testin bütününe güvenilirliği kestirilir.

Testi iki eşdeğer yarıya bölmenin bir yolu, test- teki tek numaralı sorularla çift numaralı soruları ayrı puanlamaktır.

Bu yöntemde, ölçekte yer alan sorular ikiye ayrılır ve bu iki parça arasında korelasyon hesaplanır.

Ayrıca bu iki parçadan her biri için ayrı ayrı alfa katsayıları hesaplanır.

Eğer k soru içeren ölçekteki soru sayısı çift ise, her bir alt gruba düşen soru sayısı $k/2$ 'dir.

Yani soruların ilk yarısı birinci alt gruba, ikinci yarısı ikinci alt gruba dahil edilir.

Eğer ölçekteki soru sayısı tek ise, birinci alt gruba dâhil edilecek soru sayısı $k_1=(k+1)/2$ ile bulunur. İkinci alt gruba dâhil edilecek soru sayısı ise $k_2=k-k_1$ şeklinde bulunur.

SPSS'te Split Half yöntemi ile sorular farklı biçimlerde ikiye bölünebilir.

1. Soruların ilk yarısı ve ikinci yarısı
2. Tek numaralı sorular, çift numaralı sorular
3. Rasgele
4. Kolaylık ve zorluk açısından

RELIABILITY

```
/VARIABLES=y1 y2 y3 y4 y5 y6 y7 y8 y9 y10
```

```
/SCALE('ALL VARIABLES') ALL
```

```
/MODEL=SPLIT.
```

RELIABILITY

```
/VARIABLES=y1 y3 y5 y7 y9 y2 y4 y6 y8 y10
```

```
/SCALE('ALL VARIABLES') ALL
```

```
/MODEL=SPLIT.
```

Ölçek ikiye bölümdükten sonra Spearman-Brown formülü ile güvenilirlik katsayısı bulunabilir.

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

r_{xy} = birinci yarı ile ikinci yarı arasındaki korelasyon

Araştırmacı birden fazla yarıya bölme yöntemi uygulamışsa elde edilen farklı güvenilirlik katsayılarının ortalaması alınır. Yarıya bölme güvenilirlik katsayısı en az 0,70 olmalıdır.

İkiye Bölünmüş Yöntem (Split Half) :

Reliability Statistics

Cronbach's Alpha	Part 1	Value	,836
		N of Items	5 ^a
	Part 2	Value	,796
		N of Items	5 ^b
Total N of Items			10
Correlation Between Forms			,692
Spearman-Brown Coefficient	Equal Length		,818
	Unequal Length		,818
Guttman Split-Half Coefficient			,816

a. The items are: yontem1, yontem2, yontem3, yontem4, yontem5.

b. The items are: yontem6, yontem7, yontem8, yontem9, yontem10.

Ölçek ilk 5 soru ve son 5 soru diye ikiye bölündüğünde alfa değerleri birbirine yakın ve yüksek çıkmıştır. Bu değerler soruların birbirini izleyen nitelikte düzenlendiğini ifade eder.

Bu modelde ölçeğin güvenilirliği için formlar arası korelasyonlar (Correlation Between Forms) belirler. Aynı zamanda Guttman Split Half, Eşit ve eşit olmayan uzunluk Spearman-Brown katsayılarına güvenilirlik ölçüsü olarak bakılır. Buradaki değerler ölçek güvenilirliğini yüksek olduğunu göstermektedir.

Örnek.

KATILIMCI	MADDELER						I.YARI (X)	II.YARI (Y)
	1	2	3	4	5	6		
1	4	3	2	2	3	3	9	8
2	5	4	3	4	4	3	12	11
3	2	3	2	3	3	4	7	10
4	5	5	4	2	5	4	14	11
5	2	3	2	2	1	2	7	5

$$r_{SB} = \frac{2r_{xy}}{1 + r_{xy}} = \frac{2 \times 0,69}{1 + 0,69} = 0,81$$

	s1	s2	s3	s4	s5	s6
1	4	3	2	2	3	3
2	5	4	3	4	4	3
3	2	3	2	3	3	4
4	5	5	4	2	5	4
5	2	3	2	2	1	2
6						

Reliability Analysis

Items:

- s1
- s2
- s3
- s4
- s5
- s6

Model: Split-half

Scale label:

OK, Paste, Reset, Cancel, Help

Reliability Statistics

Cronbach's Alpha	Part 1	Value	,897
		N of Items	3 ^a
	Part 2	Value	,646
		N of Items	3 ^b
	Total N of Items		6
Correlation Between Forms			,693
Spearman-Brown Coefficient	Equal Length		,818
	Unequal Length		,818

Guttman Katsayıları : Yarı testler arasındaki puan farkları kullanılır, ancak tüm testin güvenilirliğini gösteren bir indeks hesaplanır. Bu yöntemde 6 katsayı hesaplanır. Güvenirliği kovaryans ya da varyans yaklaşımı ile hesaplayan diğer bir yaklaşımdır.

Paralel Model : *Bu model ölçekteki sorular için eşit varyanslılığı ve karşılıklı tekrar soruları içinde eşit hata varyanslılığını varsayar. Bu modelle en büyük benzerlik tahminleri yapılır ve tahminlerin verilere uygunluğu ki-kare ile test edilir.*

Kesin Paralel Model (Strict Parallel): *Bu modelde eşit varyanslılığın yanı sıra sorular arasındaki ortalamalarında eşit olduğu varsayımı söz konusudur. Maddelerin ortalamalarının eşit olmadığı Hotelling T-Square test for equality of means ile test edilir. $P > 0,05$ ise soru ortalamaları arasında fark yoktur denir.*

Ölçeği oluşturan her bir sorunun tanımlayıcı istatistiklerine bakarak Güvenilirlik Analizi için hangi modelin kullanılacağına karar verilir. Sorular arasında eşit varyanslılık (homojen) varsa Alfa ve Paralel model kullanılabilir. Eşit varyanslılığın yanında sorular arasındaki ortalamalarda homojen ise Kesin Paralel model tercih edilmelidir.

Guttman Yöntemi

Reliability Statistics

Lambda	1	,794
	2	,884
	3	,882
	4	,816
	5	,856
	6	,887
N of Items		10

Guttman modeline göre hesaplanan 6 güvenilirlik katsayısının hepsi oldukça güvenilir görülmektedir.

Paralel Yöntemi

Reliability Statistics

Common Variance	1,190
True Variance	,508
Error Variance	,682
Common Inter-Item Correlation	,427
Reliability of Scale	,882
Reliability of Scale (Unbiased)	,882

Bu model ölçekteki sorular için eşit varyanslılığı ve karşılıklı tekrar soruları içinde eşit hata varyanslılığını varsayar. Bu yöntemle göre güvenilirlik katsayısı olarak Reliability of Scale değerine bakılır.

Kesin Paralel Yöntemi

Reliability Statistics

Common Mean	3,141
Common Variance	1,244
True Variance	,504
Error Variance	,740
Common Inter-Item Correlation	,404
Reliability of Scale	,871
Reliability of Scale (Unbiased)	,872

İç Tutarlılık

c). Maddeler Arası Korelasyon Katsayılarının Ortalaması

Bu analizde ölçeğin/testin toplam puanları hesaplamaya katılmaz. Burada maddeler arası korelasyon analizi yapılır ve değişkenlerin korelasyon katsayılarının ortalaması alınır. Bu analiz test/ölçek maddelerinin ne ölçüde birbirleriyle ilişkili olduğu hakkında bilgi verir.

Maddeler arasında negatif korelasyon varsa, bu maddeler arasında ters bir ilişki olduğu anlamına gelir. Negatif korelasyon olduğunda bu madde(lerin) diğer maddelerle olan ilişkisine bakılarak ölçekten çıkarılması gerekir.

İç Tutarlılık

d). Madde Toplam Puan Korelasyonu Katsayılarının Ortalaması

Bu yöntem test maddelerinden elde edilen puanlar ile testin toplam puanı arasındaki ilişkiyi açıklar. Madde toplam puan korelasyon katsayılarının ortalaması testin güvenilirliğini verir.

Madde toplam puan korelasyon katsayısı eğer 0,20'nin altında ise (örneklem eksikliğinden olabilir) bu maddelerde ciddi sorun var demektir. 0,2-0,3 çıkan korelasyonlar varsa bu maddeler eğer ciddi öneme sahipse testte kalırlar.

Madde toplam puan korelasyonu negatif çıkmışsa (yanlış kodlama, tersine çevirmeme, kasıtlı cevaplama gibi) bu maddeler ölçekten çıkarılmalıdır.

İç Tutarlılık

d). Madde Toplam Puan Korelasyonu Katsayılarının Ortalaması

Yapılan madde toplam korelasyonları için gerçekleştirilen Pearson çarpım moment korelasyon analizi sonucuna tüm maddelerle toplam puan arasındaki ilişki istatistiksel açıdan $p < 0,05$ düzeyinde pozitif yönde anlamlı bulunmalı.

Madde toplam korelasyonu, test maddelerinden alınan puanlar ile testin toplam puanı arasındaki ilişkiyi açıklamaktadır. Madde kalan korelasyonu ise ele alınan maddenin kendisi hariç diğer maddelerden elde edilen toplam puanla olan ilişkisini belirtmektedir.

Ölçmenin Standart Hatası ($SH_{\ddot{o}}$):

Testin güvenilirliği için kullanılan tekniklerden biridir. Ölçmenin standart hatası bireylerin testten aldıkları puanların gerçek puandan olan sapma miktarını açıklamada kullanılır.

$$SH_{\ddot{o}} = s \times \sqrt{1 - r}$$

R : Güvenilirlik katsayısı

s : test puanlarının standart sapması

Güven aralığı:

$$\bar{X} \mp 1,96 \times SH_{\ddot{o}}$$

KR-20 ve KR-21 Formülleri :

Bu formüller bilgi testlerinde sıklıkla kullanılır. Testteki her bir maddeye doğru cevap veren öğrenci yüzdesi hesaplanmışsa, testin güvenilirliği, Kuder-Richardson 20 ve 21 formülleriyle tahmin edilebilir. Bu yolla hesaplanan güvenilirlik katsayısı da testin iç tutarlılığının bir ölçüsüdür.

Test maddelerine verilen cevaplar 2 şıklı ise (evet/hayır, doğru/yanlış gibi) KR-20 ve KR-21 formülleri kullanılabilir. Test maddelerinin aynı ortalama ve varyansa sahip olduğu varsayımı ile hareket eder. Dolayısıyla da maddeler arası korelasyonlara bakılır. Maddeler iki şıklı puanlanmış ise bir madde (soru) nin doğru cevaplanma olasılığı (p) ve yanlış cevaplanma olasılığı (q) hesaplanabilir.

KR-20 formülü, test maddelerinin benzer zorlukta olmadığı varsayımı altında kullanılır. KR-20 formülünde temel varsayım ölçü maddelerinin tek bir yapıyı ölçüyor olmasıdır. Maddelerin içerikleri benzer olmalıdır. Maddeler iki kategorili olduğunda KR-20 ile Cronbach alfa aynı sonucu verir.

KR-20 Formülü

$$KR - 20 = \frac{K}{K - 1} \left[1 - \frac{\sum pq}{s^2} \right]$$

p : Soruların (maddelerin) doğru cevap oranı

q : Soruların (maddelerin) yanlış cevap oranı

K : Testteki soru (madde) sayısı

s^2 : Toplam puanların (herbir kişinin doğru cevap sayısı) varyansı (varyans bulunurken n-1 yerine n kullanılır).

KR-20>0,70 ise testin iç güvenilirliği yüksek demektir.

Örnek.

Öğ\SORU	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	T
1	1	0	0	1	1	1	0	1	1	1	7
2	1	1	0	1	1	0	1	1	1	1	8
3	0	0	1	1	1	0	0	1	0	1	5
4	0	1	0	1	0	1	1	1	1	1	7
5	0	0	0	0	0	0	0	1	0	0	1
6	1	0	0	0	0	1	1	1	0	0	4
7	1	0	0	0	1	0	0	0	1	1	4
8	1	0	1	0	0	1	1	0	1	1	6
9	1	1	1	1	1	0	0	1	1	1	8
10	1	1	1	0	0	1	1	1	1	1	8
p	0,7	0,4	0,4	0,5	0,5	0,5	0,5	0,8	0,7	0,8	
q	0,3	0,6	0,6	0,5	0,5	0,5	0,5	0,2	0,3	0,2	
pq	0,21	0,24	0,24	0,25	0,25	0,25	0,25	0,16	0,21	0,16	2,22

$$KR - 20 = \frac{k}{k-1} \left[1 - \frac{\sum pq}{s^2} \right] = \frac{10}{10-1} \left[1 - \frac{2,22}{4,76} \right] = 0,593$$

Soru.

Öğ\ soru	S1	S2	S3	S4	S5	S6	T
1	1	1	1	1	0	1	5
2	0	0	1	0	1	0	2
3	0	0	0	1	0	1	2
4	0	0	1	1	1	0	3
5	1	1	1	1	0	0	4
6	0	1	1	0	1	0	3
7	0	1	0	0	0	0	1
8	1	1	1	1	1	1	6
9	0	1	1	0	1	0	3
10	1	1	1	0	0	1	4
p	0,4	0,7	0,8	0,5	0,5	0,4	
q	0,6	0,3	0,2	0,5	0,5	0,6	
pq	0,24	0,21	0,16	0,25	0,25	0,24	1,35

$$KR - 20 = \frac{k}{k-1} \left[1 - \frac{\sum pq}{s^2} \right] = \frac{6}{6-1} \left[1 - \frac{1,35}{2,23} \right] = 0,473$$

KR-21 formülü eşit zorluğa sahip maddelerden oluşan (her sorunun güçlük derecesi aynı ise) ve sınıf ortamında uygulanan çoktan seçmeli testler için uygundur. Doğru cevap 1, yanlış cevaplar ise 0 olarak kodlanır. Testteki soruların zorluk dereceleri eşit olmalıdır.

$$KR - 21 = \frac{k}{k-1} \left[1 - \frac{\bar{x}(k - \bar{x})}{ks^2} \right]$$

k: Madde (soru) sayısı

KR-20 ve KR-21 formüllerinden elde edilen katsayı 0,7'nin üstünde ise güvenilirlik iyidir. KR-20 ve KR-21 formülleri testin genel olarak güvenilirliği hakkında bilgi verir, tek tek maddelerin güvenilirliği hakkında bilgi vermez.

Soru.

Öğ\ soru	S1	S2	S3	S4	S5	S6	T
1	1	1	1	1	0	1	5
2	0	0	1	0	1	0	2
3	0	0	0	1	0	1	2
4	0	0	1	1	1	0	3
5	1	1	1	1	0	0	4
6	0	1	1	0	1	0	3
7	0	1	0	0	0	0	1
8	1	1	1	1	1	1	6
9	0	1	1	0	1	0	3
10	1	1	1	0	0	1	4
p	0,4	0,7	0,8	0,5	0,5	0,4	
q	0,6	0,3	0,2	0,5	0,5	0,6	
pq	0,24	0,21	0,16	0,25	0,25	0,24	1,35

$$KR - 21 = \frac{k}{k-1} \left[1 - \frac{\bar{x}(k - \bar{x})}{ks^2} \right] = \frac{6}{6-1} \left[1 - \frac{3,3(6-3,3)}{6 \times 2,01} \right] = 0,313$$

f) Cronbach Alfa Katsayısı:

Alfa katsayısı ölçekte yer alan k sorunun varyansları toplamının genel varyansa oranlaması ile bulunan bir ağırlık standart değişim ortalamasıdır. Sorular arasında negatif korelasyon varsa Alfa katsayısı da negatif çıkar. Bu durum güvenilirlik modelin bozulmasına neden olur.

Cronbach Alfa Katsayısı, ölçekte yer alan k sorunun homojen bir yapıyı açıklamak üzere bir bütün oluşturup oluşturmadıklarını araştırır.

Alfa Katsayısı, bireysel puanların k soru içeren bir ölçekte sorulara verilen cevapların toplanması ile bulunduğu durumlarda soruların birbirleri ile benzerliğini, yakınlığını, ortaya koyan bir katsayıdır.

Alfa ne kadar yüksek olursa ölçekteki maddelerin birbiriyle tutarlı ve aynı özelliği taşıyan maddeler olduğu anlaşılır.

- Cronbach alfa katsayısı istatistik temelleri tutarlı ve tüm soruları dikkate alarak hesaplandığından, genel güvenilirlik yapısını diğer katsayılara göre en iyi yansıtan katsayıdır.
- Sorular arası korelasyonu hesaplamanın bir yolu da Cronbach ın alfasını hesaplamaktır. Eğer formdaki sorular güvenilir ise bu katsayı 1 yakın olur. Cronbach alfa test-tekrar testle hesaplanan güvenilirlik katsayısından daha az konservatiftir.

- Güvenirlilik Analizi, toplam puanlar üzerine kurulu ölçeklere (Likert Ölçeği, Q-Tipi Ölçek) dayalı araçların güvenirliliğini ortaya koymaya yarayan Cronbach Alfa katsayıları hesaplar.

Temel varsayımlar,

- "Her soru toplam skorun bir doğrusal bileşeni olmalıdır." ve
- "Ölçekte toplanabilirlik özelliğinin bulunması gerekir." biçimindedir.

- **Cronbach Alfa katsayısı**, bireysel puanların k soru içeren bir ölçekte sorulara verilen cevapların toplanması ile bulunduğu durumlarda soruların birbirleri ile benzerliğini, yakınlığını ortaya koyan bir katsayıdır.
- Alfa katsayısı, ölçekte yer alan k sorunun türdeş bir yapıyı açıklamak ya da sorgulamak üzere bir bütün oluşturup oluşturmadıklarını sorgulamayı sağlar.

Cronbach Alfa deęişik biçimlerde hesaplanabilir:

1. Maddelerin varyans deęerlerinden hareketle hesaplama

$$\alpha = \frac{k}{k-1} \left[\frac{\sigma_t^2 - \sum \sigma_i^2}{\sigma_t^2} \right]$$

σ_t^2 = Toplam Sütunun Varyansı

σ_i^2 = Deęişkenlerin her birinin varyansı

Tablo 2. Alfa Katsayısının Hesaplanması

Kişiler	Madde 1	Madde 2	Madde 3	Toplam
1	2	3	3	8
2	3	4	3	10
3	3	2	2	7
4	2	4	4	10
5	4	4	1	9
<i>O ortalama</i>	2,8	3,4	2,6	8,8
<i>SS Standart Sapma</i>	0,84	0,89	1,14	1,30
<i>S²Varyans</i>	0,7	0,8	1,3	1,7

$$\alpha = \frac{3}{2} \left[\frac{1,7 - 2,8}{1,7} \right] = -0,97$$

2. Maddelerin kovaryans değerlerinden hareketle hesaplama

$$\alpha = \frac{\frac{k \times Ort(kov)}{Ort(Var)}}{1 + \frac{(k - 1) \times Ort(kov)}{Ort(Var)}}$$

k: ölçekte yer alan madde sayısı

Ort(kov): Ortalama kovaryans

Ort(Var): Ortalama varyans

Standardized Cronbach's Alpha

Korelasyon matrisi verilerinden hareketle hesaplama:

$$\alpha = \frac{k\bar{r}}{1 + (k - 1)\bar{r}}$$

k: ölçekte yer alan madde sayısı

\bar{r} : maddeler arasındaki ortalama korelasyon katsayısı

Güvenilirlik Analizi Büyüklüğünü Etkileyen Faktörler

- Güvenirlik analizinde k soru ve n birim vardır. Güvenirlik analizleri yapılırken k ve n ye ait iki temel koşulun dikkate alınması gerekir;

i) $K > 20$ olmalıdır. Yani ölçme araçları , bireysel özellikleri (demografik, sosyo ekonomik) sorgulayan sorular dışında, çok sayıda ve birbirleri ile ilişkili soru içermelidir.

ii) $n > 50$ olmalıdır. Yani ölçme araçları çok sayıda rasgele seçilen deneklere uygulanmalıdır.

- Ölçekteki gözlem sayısı, soru (madde) sayısından fazla olmalıdır.
- Madde sayısı ve n arttıkça testin güvenilirliği artabilir.
- Bireyler test edilmeye hazır olmalıdırlar. Bunun için bireyler testten önce test hakkında bilgilendirilmelidirler.
- Bireylerin yaş, cinsiyet, eğitim ve deneyim düzeyleri güvenilirlik katsayısının büyüklüğünün etkileyebilir.
- Testi yapan kişinin özellikleri önemlidir. Testi yapan kişi deneyimli ve objektif ise güvenilirlikte yüksek olacaktır.
- Test özellikleri güvenilirlik katsayısının büyüklüğünü etkiler. Örneğin fiziksel uygunluk testleri spor beceri testlerinden daha güvenilirdir. Çünkü fiziksel uygunluk testleri daha objektif iken, spor beceri testleri daha subjektif testlerdir.

Güvenirlilik analizinin temel varsayımları:

- Her soru toplam skorun bir doğrusal bileşeni olmalıdır.
- Ölçekte toplanabilirlik özelliğinin bulunması gerekir.

Cronbach Alfa Katsayısı

- Sorular arasında negatif korelasyon varsa Cronbach Alfa Katsayısı da negatif çıkar.
- Alfa'nın negatif çıkması, güvenirlik modelinin bozulmasına neden olur.
- Çünkü bu durumda ölçeğin toplanabilirlik varsayımı bozulmuş ve ölçek toplanabilir ölçek olmaktan çıkmış olur.

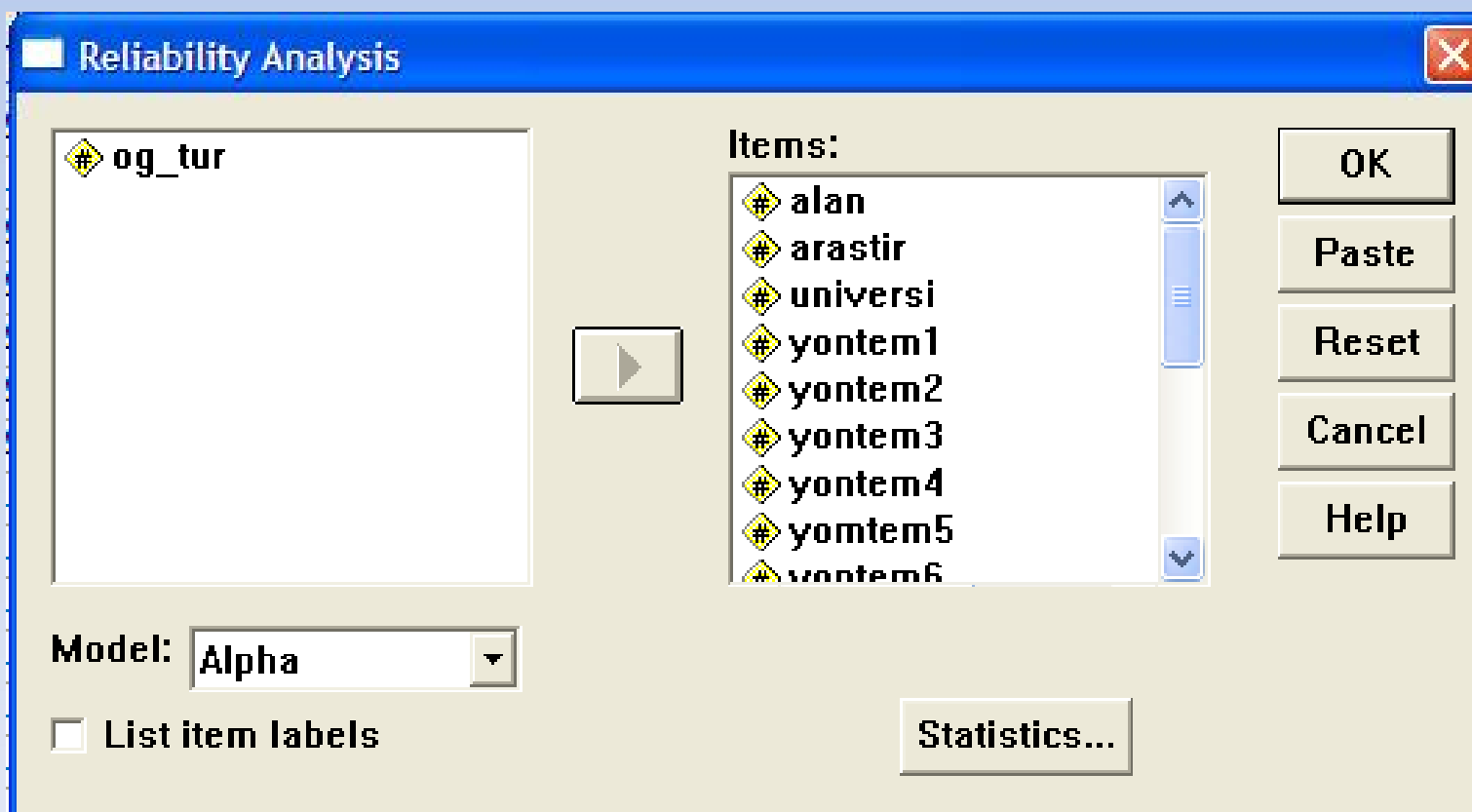
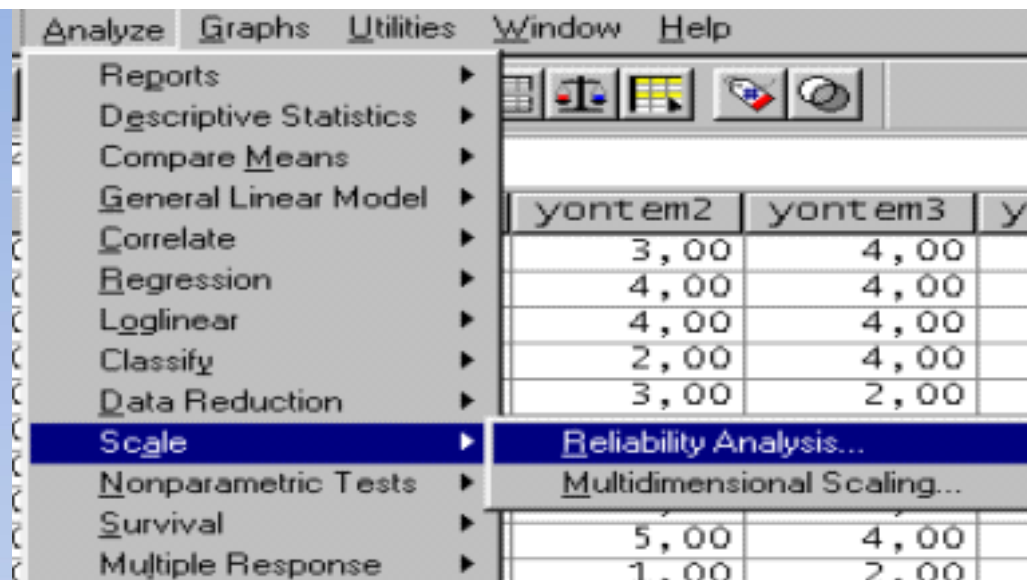
Alfa'nın negatif çıkmasının nedenleri:

- Bireysel maddelere ait varyansların toplamının toplam puanlara ait varyanstan büyük olması
- Araştırmada ciddi ölçüm hatalarının olması
- Ters yönlü maddelerin kodlanmasına dikkat edilmemesi
- Örneklem hacmi küçük veya ölçekteki madde sayısının az olması
- Maddeler arası kovaryans değerlerinin ortalamasının negatif olması
- Testin birden fazla faktör içermesi ve bu faktörlerin birbiriyle negatif yönde ilişkili olması

- Alfa katsayısının bulunabileceği aralıklar ve buna bağlı olarak da ölçeğin güvenilirlik durumu aşağıda verilmiştir
- $0,00 \leq \alpha < 0,40$ ise ölçek güvenilir değildir,
- $0,40 \leq \alpha < 0,60$ ise ölçek düşük güvenilirliktedir,
- $0,60 \leq \alpha < 0,80$ ise ölçek oldukça güvenilirdir,
- $0,80 \leq \alpha < 1,00$ ise ölçek yüksek derecede güvenilir bir ölçektir.

- ❖ Bir ölçekteki sorulara verilen cevapların birey ve sorulara göre önemliliğini belirlemek için iki yönlü varyans analizi yapılır.
- ❖ Sorular arasındaki benzerlikler F testi ile analiz edilir.
- ❖ Sorular sıralama puanları olarak alınmış ise, birey ve soru farklılıklarının analizi için Friedman Ki-Kare testi kullanılır.
- ❖ Sorulara verilen cevaplar 0, 1 biçiminde ikili tipte iseler önemlilik değerlendirilmesi Cochran Ki-Kare testi yardımı ile yapılır.
- ❖ Ölçeğin toplanabilir ölçek tipinde hazırlanıp hazırlanmadığını test etmek için Tukey eklenebilirlik testi (Tukey's Additivity Test) kullanılır.
- ❖ Soru ortalamalarının birbirlerine eşit olup olmadıkları ise Hotelling T² istatistiği ile test edilir.

SPSS'TE CRONBACH ALFA HESABI



Reliability Analysis: Statistics

Descriptives for

- Item
- Scale
- Scale if item deleted

Inter-Item

- Correlations
- Covariances

Summaries

- Means
- Variances
- Covariances
- Correlations

ANOVA Table

- None
- F test
- Friedman chi-square
- Cochran chi-square

Hotelling's T-square

Tukey's test of additivity

Intraclass correlation coefficient

Model: Type:

Confidence % Test value:

Continue
Cancel
Help

Item : soru

Scale : ölçek

F test: Ölçeği oluşturan sorular arasındaki benzerlik analizini yapar.

Friedman chi-square: Ölçekteki sorulara verilen cevaplar sıralama puanları olarak verilmişse birey ve soru farklılıklarının analizi yapılır.

Cochran chi-square: Ölçekteki sorulara verilen cevaplar 0 yada 1 biçiminde iki değerli ise birey ve sorulara göre önemlilik analizi yapılır.

Tukey's test of additivity : Ölçekte yer alan soruların bir toplamsal ölçek oluşturacak biçimde hazırlanıp hazırlanmadığını, soruların skorlarının toplanabilir olup olmadığını test eder.

Interclass correlation coefficient: Değerlendiriciler arasındaki tutarlılığı ve uyumları test eder ve hesaplama modeli için iki yönlü rasgele (Two-way Random) model seçilir. Birimler içi değerlerin tutarlılığını ve mutlak uyumlarını ölçen katsayıları hesaplar.

Hotelling's T-square : Soru ortalamalarının birbirine eşit olup olmadığını test eder. Soru ortalamalarının birbirine eşit olup olmaması kavramı; soruların denekler tarafından aynı yaklaşım ile algılanıp algılanmadığını yani soruların zorluk derecelerinin birbirine eşit olup olmadığıdır. Ölçek değerlendirmede soruların %75'i orta, %12,5'i kolay ve %12,5'i de zor nitelikte olmalıdır. sıfır olarak seçilir.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,882	,886	10

Alfa



Item Statistics

	Mean	Std. Deviation	N
yontem1	3,24	,973	348
yontem2	3,42	,979	348
yontem3	3,35	,944	348
yontem4	3,29	1,128	348
yomtem5	3,04	1,039	348
yontem6	3,11	1,037	348
yontem7	3,20	1,086	348
yontem8	2,69	1,085	348
yontem9	2,76	1,316	348
yontem10	3,31	1,262	348

Inter-Item Correlation Matrix

	yontem1	yontem2	yontem3	yontem4	yontem5	yontem6	yontem7	yontem8	yontem9	yontem10
yontem1	1,000	,559	,521	,445	,426	,447	,455	,345	,203	,294
yontem2	,559	1,000	,568	,554	,407	,449	,508	,402	,246	,305
yontem3	,521	,568	1,000	,560	,517	,456	,511	,416	,322	,340
yontem4	,445	,554	,560	1,000	,518	,560	,502	,396	,292	,447
yontem5	,426	,407	,517	,518	1,000	,573	,529	,367	,400	,405
yontem6	,447	,449	,456	,560	,573	1,000	,609	,393	,351	,327
yontem7	,455	,508	,511	,502	,529	,609	1,000	,571	,436	,436
yontem8	,345	,402	,416	,396	,367	,393	,571	1,000	,348	,500
yontem9	,203	,246	,322	,292	,400	,351	,436	,348	1,000	,487
yontem10	,294	,305	,340	,447	,405	,327	,436	,500	,487	1,000

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	3,141	2,690	3,417	,727	1,270	,060	10
Item Variances	1,190	,891	1,731	,840	1,942	,077	10
Inter-Item Covariances	,508	,260	,809	,549	3,108	,012	10
Inter-Item Correlations	,438	,203	,609	,406	2,996	,009	10

Ölçekte yer alan soruların genel ortalaması (item means) 3,141'dir. Ortalanmaların değişim aralığı (range) 0,727 olmaktadır. Soru-Bütün (Inter-item) korelasyonlara bakıldığında 0,203 ile 0,609 arasında değişim gösterdiği görülür. Soru ile bütün arasındaki korelasyonların negatif olmaması gerekir, hatta 0,2'den büyük olması beklenir.

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
31,41	57,661	7,593	10

10 sorudan oluşan ölçek ortalaması (toplam puan) 31,41 ve standart sapma 7,593 bulunmuştur.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
yontem1	28,18	49,017	,565	,412	,874
yontem2	28,00	48,303	,617	,488	,870
yontem3	28,06	48,152	,658	,495	,868
yontem4	28,13	46,134	,669	,518	,866
yomtem5	28,37	47,254	,653	,469	,867
yontem6	28,30	47,259	,654	,513	,867
yontem7	28,22	45,812	,726	,568	,862
yontem8	28,72	47,659	,589	,420	,872
yontem9	28,65	47,197	,483	,329	,883
yontem10	28,10	46,408	,562	,420	,875

Scale Mean/Variance if Item Deleted: Ölçekten bir soru çıkartıldığında geriye kalan soruların ölçek ortalaması/varyansı

Corrected Item-Total Correlation : Ölçekten çıkarılan ilgili soru ile ölçekteki toplam puan arasındaki korelasyon (Point Biserial Korelasyon). Bu değerler negatif olmamalı ve 0,20'den büyük olmalıdır.

Squared Multiple Correlation : Ölçekten ilgili soru çıkartıldıktan sonra kalan kısma ait çoklu korelasyon katsayısı

Cronbach's Alpha if Item Deleted: Bir sorunun ölçekten çıkarılması için; soru silinirse Alfa katsayısında (Alpha if Item Deleted) değişimine bakılır. Bu sütunda en yüksek değerli soru ölçekten çıkarıldığında, eğer alfa katsayısı yükseliyorsa o soru güvenilirliği azaltan bir sorudur ve ölçekten çıkarılması gerekir.

Madde Silindiğinde Ortalama-Varyanslardaki Değişim (Scale Means-Variance if item deleted)

Eğer madde ölçekten çıkarılırsa bireylerin madde ortalamalar ve varyansda büyük değişiklik olmaması istenir.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
yontem1	28,18	49,017	,565	,412	,874
yontem2	28,00	48,303	,617	,488	,870
yontem3	28,06	48,152	,658	,495	,868
yontem4	28,13	46,134	,669	,518	,866
yontem5	28,37	47,254	,653	,469	,867
yontem6	28,30	47,259	,654	,513	,867
yontem7	28,22	45,812	,726	,568	,862
yontem8	28,72	47,659	,589	,420	,872
yontem9	28,65	47,197	,483	,329	,883
yontem10	28,10	46,408	,562	,420	,875

Madde Bütün Korelasyon Katsayısı (Corrected item-total correlation)

Ölçekteki her bir maddenin o madde dışındaki maddelerin toplamıyla-bütünüyle olan korelasyon katsayısıdır. Bu Point Biserial Korelasyondur.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
yontem1	28,18	49,017	,565	,412	,874
yontem2	28,00	48,303	,617	,488	,870
yontem3	28,06	48,152	,658	,495	,868
yontem4	28,13	46,134	,669	,518	,866
yontem5	28,37	47,254	,653	,469	,867
yontem6	28,30	47,259	,654	,513	,867
yontem7	28,22	45,812	,726	,568	,862
yontem8	28,72	47,659	,589	,420	,872
yontem9	28,65	47,197	,483	,329	,883
yontem10	28,10	46,408	,562	,420	,875

Çoklu Açıklayıcılık Katsayısı (R^2) (Squared Multiple Coreelation)

Bir madde ile geriye kalan diğer maddeler arasındaki ilişki R^2 ile bulunabilir. Bunun için ilgili madde bağımlı, diğer maddeler bağımsız değişken alınarak çoklu doğrusal regresyon analizi yapılarak R^2 bulunabilir. Bir maddenin R^2 değerinin küçük çıkması, o maddenin diğer maddeler tarafından yeterince açıklanmadığını gösterir.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
yontem1	28,18	49,017	,565	,412	,874
yontem2	28,00	48,303	,617	,488	,870
yontem3	28,06	48,152	,658	,495	,868
yontem4	28,13	46,134	,669	,518	,866
yontem5	28,37	47,254	,653	,469	,867
yontem6	28,30	47,259	,654	,513	,867
yontem7	28,22	45,812	,726	,568	,862
yontem8	28,72	47,659	,589	,420	,872
yontem9	28,65	47,197	,483	,329	,883
yontem10	28,10	46,408	,562	,420	,875

Madde Silinirse Güvenilirlik Katsayısı (Cronbach's alpha if item deleted)

Ele alınan madde ölçekten çıkarıldığında güvenilirlik katsayısının değişimini incelemek amacıyla yararlanılan bir yaklaşımdır. Bir madde ölçekten çıkartıldığında alfa katsayısı ölçeğin tümü için hesaplanan alfa katsayısına göre artış gösteriyorsa, o madde ölçeğin güvenilirliğini azaltan bir maddedir. Eğer bir madde ölçekten çıkarıldığında alfa katsayısı ölçeğin tümü için hesaplanan alfa katsayısından küçük çıkıyorsa o madde ölçekte yer almalıdır. Böyle bir madde olup olmadığına bakmak için **Cronbach's alpha if item deleted** değerlerinden en büyük olana bakılır.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
yontem1	28,18	49,017	,565	,412	,874
yontem2	28,00	48,303	,617	,488	,870
yontem3	28,06	48,152	,658	,495	,868
yontem4	28,13	46,134	,669	,518	,866
yontem5	28,37	47,254	,653	,469	,867
yontem6	28,30	47,259	,654	,513	,867
yontem7	28,22	45,812	,726	,568	,862
yontem8	28,72	47,659	,589	,420	,872
yontem9	28,65	47,197	,483	,329	,883
yontem10	28,10	46,408	,562	,420	,875

ANOVA with Tukey's Test for Nonadditivity

			Sum of Squares	df	Mean Square	F	Sig
Between People			2000,841	347	5,766		
Within People	Between Items		187,591	9	20,843	30,560	,000
	Residual	Nonadditivity	,329 ^a	1	,329	,482	,488
		Balance	2129,681	3122	,682		
		Total	2130,009	3123	,682		
Total			2317,600	3132	,740		
Total			4318,441	3479	1,241		

Grand Mean = 3,14

a. Tukey's estimate of power to which observations must be raised to achieve additivity = 1,173.

Yapılan varyans analizi sonucuna göre ölçek toplanabilir özelliktedir (Nonadditivity: $F=0,482$ $P=0,488>0,05$). Ölçümler arası değişime bakıldığında önemli bir farklılık görülmektedir (Between Measures, $F=30,56$ $P=0,00<0,05$).

Yani 10 soruluk bu ölçek toplanabilir özelliktedir ancak ölçümler arasında farklılıklar vardır.

Hotelling's T-Squared Test

Hotelling's T-Squared	F	df1	df2	Sig
282,420	30,657	9	339	,000

Soru ortalamalarının eşitliğini test eden Hotelling T² testi sonucunda, soru ortalamaların farklı olduğu görülmektedir (Hotelling's T-Squared=282,42 , P=0,000<0,05). Yani en az iki ortalama arasında farklılık vardır. Bu farklılıkların hangi sorulardan kaynaklandığı araştırılmalıdır.

Intraclass Correlation Coefficient

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	,427 ^b	,385	,472	8,454	347	3123	,000
Average Measures	,882	,862	,899	8,454	347	3123	,000

Two-way random effects model where both people effects and measures effects are random.

- Type C intraclass correlation coefficients using a consistency definition-the between-measure variance is excluded from the denominator variance.
- The estimator is the same, whether the interaction effect is present or not.

Intraclass Correlation (Küme-sınıf içi korelasyon analizi): Gözlemci (hakem) değerlendirmelerinin güvenilirliğini belirlemede kullanılır. Bir kişiye ait gözlem değerlerinin diğer kişilerin gözlem değerlerine ne ölçüde benzer olduğunu gösterir.

Single Measures Intraclass Correlation : Tek ölçüm küme (sınıf) içi korelasyon katsayısı

Average Measures Intraclass Correlation : Rasgele seçilen kişiler çok sayıda farklı gözlemciler(hakemler) tarafından değerlendirilir ve bu değerlendirmelerin ortalaması alınarak korelasyon değeri hesaplanır.

F testi ile ölçüğü oluşturan sorular arasında benzerlik olup olmadığı test edilmiş ve benzerlik olmadığı görülmüştür (F=8,454, p=0,000<0,05).

Soru (Madde) ile Bütün Arasındaki Korelasyonlar Yöntemi (Item Total Correlation)

Bir madde ile bu madde dışında diğerk maddelerin toplamından oluşan bütün (total) arasındaki korelasyon hesaplanmasına dayanmaktadır. Eğer madde-bütün korelasyon katsayısı düşük ise, o sorunun ölçeğe katkısının düşük olduğu anlamına gelir ve o maddenin ölçekten çıkarılması gerekir.

Madde-bütün korelasyon katsayısının eksi işaretli olmaması ve $>0,25$ olması istenir.

Madde-bütün korelasyon katsayısı maddenin ayırdediciliği ya da maddenin geçerliliği konusunda bilgi verir.

Madde-bütün korelasyon katsayısı hesaplanırken dikkatli olunmalıdır. Özellikle 5'li ve 7'li Likert tipi ölçeklerde seçenekler arasındaki farkın eşit ve bütün puanın sürekli olduğu varsayımı altında, madde-bütün korelasyonları genellikle Pearson korelasyon katsayısı ile hesaplanır. Eğer madde seçenekleri iki kategorili (başarılı-başarısız, evet-hayır, var-yok, katılıyorum-katılmıyorum vb.) ise madde-bütün korelasyonu nokta-iki serili korelasyon (r_{pb}) ile hesaplanmalıdır.