

ANKET, GÜVENİLİRLİK -GEÇERLİLİK ANALİZİ



PROF.DR.YÜKSEL TERZİ

ONDOKUZ MAYIS ÜNİVERSİTESİ
FEN-EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ
SAMSUN
2019

GEÇERLİLİK (Validity)

Geçerlilik;

- *Kullanılan ölçüm aracının ölçülmek istenen özelliğe uygun olması,*
- *Ölçümün kurallara uygun olarak doğru yapıp yapılmadığı,*
- *Ölçüm verilerinin gerçekten ölçülmek istenen özelliği yansıtıp yansıtmasındadır.*

Geçerlilik, belli bir olguya ait ölçüm rakamları olguyu doğru bir şekilde yansıtıyor, tanımlıyor veya kuramsal açıklamalar getiriyorsa geçerlidir (Hammersley, 1987).

Ölçüm verilerinin doğrulamasını yapmak için önce güvenilirlik analizi yapılır, ancak bu tek başına yeterli değildir. Aynı zamanda geçerlilik analizleri de yapılmalıdır.

Yüksek geçerlilik aynı zamanda yüksek güvenilirlik anlamına gelebilir. Ancak yüksek güvenilirlik geçerlilik hakkında hiçbir bilgi vermez.

Güvenilirlik analizlerinde teknik hesaplamalar ön plana çıkarken, geçerlilik analizlerinde yargısal değerlendirmeler ile teknik hesaplamaların birlikte kullanılması gerekir.

Geçerlilik bir ölçme aracının ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan, tam ve aynı zamanda doğru olarak ölçmesidir. Geçerlik bir gerecin ölçmesi için düzenlendiği olguyu ölçme derecesiyle ilişkilidir. Geçerlik sayesinde ölçülmek istenen değişkenin ölçülebilme derecesi ortaya konulmuş olur (Ergin, 1995 ; Turgut, 1989; Yılmaz, 1996).

Geçerlik, test değerlendirmede göz önüne alınması gereken en önemli konulardan bir tanesidir. Test puanlarına dayalı olarak varılabilecek çıkarımların uygunluğu, anlamlılığı ve kullanılışlılığı hakkında bilgi verir. Yani kavram, bir teste dayalı çıkarsamaların uygunluk veya anlamlılık derecesidir. Testin geçerli hale getirilmesi, bu tür çıkarsamaları desteklemek için veri toplama sürecidir (American Psychological Association, 1992).

İdeal bir geçerlilik çalışması, çeşitli kategorideki bilgilerin bir araya getirilmesini gerektirmektedir (American Psychological Association, 1992). Geçerlik çalışması için veriler çeşitli yollarla toplanabilir. Geleneksel olarak geçerlik türleri kapsamlı bağlantılı, kriterle bağlantılı veya yapıyla bağlantılı geçerlik olarak gruplanmaktadır. Bu kategoriler arasında kesin ayırım yapmak mümkün değildir. İdeal olan bir geçerlik saptama bu üç geleneksel kategoriye de kapsayan bilgileri içermelidir (Hovardalıoğlu, Sezgin, 1998; Turgut, 1989; Yılmaz, 1996).

Geçerlilikte Temel Esaslar

Geçerlilik kavramı içerisinde genelde üzerinde durulan hususlar özetlenecek olursa 3 temel esas dikkati çekmektedir.

1. Kullanılan ölçüm aracı, ölçülmek istenen özelliğe uygun bir araç mıdır?
2. Ölçüm, kurallara uygun olarak yapılıyor mu?
3. Ölçülen veriler gerçekten ölçmek istediğimiz özelliği yansıtıyor mu?

- ✓ İe kapanıklığı len bir test ie kapanıklığı, kaygıyı len bir test kaygıyı lmelidir. rneğın gen sporcuların ileride sprinter olup olmayacağını belirlemek amacıyla bir test geliştirilmiş ise bu test istenen amacı sağlamalıdır. Yanlıřlıkla orta ya da uzun mesafeci olacakları sprinter grubuna katmamalıdır.
- ✓ Geometri dersi genel sınavında ğrencilerin geometri becerileri lülecekse, sınavda geometriden hari matematik sorusu sorulmamalıdır.

Geçerlilik:

Ölçümlerde güvenilirlik tek başına yeterli değildir. Psikolojik değerlendirme testi “prenology” geçmiş dönemde kişilik ve davranış ölçümlerinin değerlendirilmesinde de kullanılıyordu.

Bu testi kaç psikolog uygularsa uygulasin aynı sonucu buluyordu. Bu ölçüm depresyonun bir göstergesi olarak kullanılırsa oldukça güvenilir bir ölçüm olabilir ama bunu kişilik veya davranışın dolaylı bir göstergesi olarak kullanmak psikolog’u yanlış yere götürür.

Yani test geliştirildiği amaç dışında bir kavramın (davranış) değerlendirmesi için bu ölçüm **geçerli** bir ölçüm değildir.

- Bu üniversite sınavlarında da çoğu zaman tartışma konusu olmaktadır. Üniversite sınavlarının güvenilir olması için, aynı sorular sorulduğunda aynı öğrencinin birbirine çok yakın sonuçlar alması beklenir. Amaç bilgiyi ölçmek ise yapılan üniversite sınavının gerçek bilgiyi ölçtüğü söylenebilir mi? Yani bu sınavın bilginin ötesinde bir takım farklı değerlendirmeleri de içeren bir sınav olduğu her zaman tartışılmıştır. İşte burada itiraz konusu olan güvenilirlik değil **geçerlilik**dir.
- Sözlü sınavlarda öğretmenin kişiler hakkındaki ön kanaatleri güvenirliliği değil geçerliliği zedelemektedir. Çünkü o öğretmen kendi sempatisinin yarattığı puanı bir kişiye ekleyecekse o sınavı 3 kere de yapsa sempati nedeniyle daha kolay sorular soracak ve verilen puan her zaman pozitif yönde etkilenecektir.

Belli bir nesneye ait bir özellik ölçüldüğünde bu ölçümden elde edilen veriler nesnenin ölçülen özelliğini doğru yansıtıyor mu, o özelliğin gerçek değerini tanımlıyor veya bu özelliğe kavramsal bir açıklama getiriyor mu sorusu geçerliliğin konusudur.

Burada geçerli olan ölçme aracı değil onunla elde edilen sonuçlarla ilgili bir husustur. Yani sonuçlar geçerlidir veya değildir şeklinde karar verilir. Söz konusu ölçme aracının geliştirildiği amaç için kullanılması ile ancak geçerli sonuçlar elde edilebilir. Metreyi ağırlık ölçmek için kullanırsak geçerli sonuç elde edemeyeceğimiz açıktır.

Geçerlilik bir test veya ölçüğün ölçülmek istenen şeyi ölçme derecesidir.

Örneğin biz cebirsel işlemleri (toplama, çıkarma, çarpma, bölme) öğrettik ve bunların öğrenilip öğrenilmediğini kontrol için bir test uyguluyoruz. Test soruların bu konular kapsamaması gerekir. Bilgisayarlar nasıl toplama yapar? Şeklinde bir soru kapsam dışı bir soru olacaktır. Testin geçerliliği zedelenecektir.

GEÇERLİLİK ANALİZ YÖNTEMLERİ

- 1. Yüzey Geçerliliği***
- 2. Kapsam(İçerik) Geçerliliği***
- 3. Kriter Geçerliliği***
- 4. Yapı Geçerliliği***

GEÇERLİLİK ANALİZ YÖNTEMLERİ

1. *Yüzey (Görünüş) Geçerliliği (Face Validity)*

Ölçme aracının neyi ölçtüğü değil de neyi ölçer göründüğünü belirtmektedir. Bir ölçeğin görünüş geçerliliği o ölçeğin ölçmek istediği özelliği ölçüyor gözükmesidir.

Ölçek Türkçeye uzman kişilerce çevrilir ve kontrol edilir. Daha sonra Türkçeye çevrilen ölçek deneme amacıyla ön çalışmada belli sayıda kişi üzerinde uygulanır ve ölçeğin anlaşılabilirliği test edilir. Pilot çalışmada deneklerin anlamada güçlük çektikleri sorular yeniden düzenlenir ve ölçek daha kolay anlaşılır hale getirilir. Türkçe form ve İngilizce form aynı kişilere farklı zamanlarda doldurulur. Sonuçlar paired t-test ile test edilir. Tüm sorularda $p > 0,05$ çıkarsa İngilizce ve Türkçe form a verilen cevaplar aynıdır sonucuna ulaşılır.

Bir testin/ölçeğin araştırılan yapıyı ölçüp ölçmediğine ilişkin olarak arařtırmacının

- i.) Kendisinin,**
- ii.) Yakın çevresindeki arkadaşlarının,**
- iii.) Arařtırılan konu hakkındaki uzman olmayan diđer kiřilerin,**
- iv.) Pilot arařtırmaya katılan cevaplayıcıların kanaat ve görüşlerinin toplanmasıyla belirlenir.**

Yüzey geçerliliğinde ölçek maddelerinin düzgün ve anlamlı bir şekilde ifade edilmesi, doğru terimlerin kullanılması, uygun kelimelerin seçilmesi, kelimelerin anlamı açık ve net olması, belirsiz birden fazla anlama gelen kelimelerden kaçınılması önemlidir.

Kullanılan test/ölçek cevaplayıcıların eğitim düzeylerine, bilgisine, kültürel ve yetenek düzeylerine uygun olmalıdır.

Cevaplayıcıların kendilerini mağdur olmuş göstermek istemeleri, mutlu ve huzurlu olduklarını yansıtmaya çalışmaları, arařtırmacının beklediđi cevapları vermemeleri ölçüm hatalarına sebep olmaktadır.

2. Kapsam (İçerik) Geçerliliği (Content validity)

Kapsam geçerliliği örneklem olarak belirlenen test/ölçek maddelerinin belirli bir amaca yönelik olarak kavramsal ana kütle temsil etme derecesidir.

Geliştirilen test incelenen konuların tüm önemli alt konularını içeriyorsa, testin kapsam-içerik geçerliliğinin olduğu söylenir.

Kapsam geçerliliği bir bütün olarak ölçeğin ve ölçekteki her bir maddenin amaca ne derece hizmet ettiği için.

Kapsam geçerliliğinde ölçüm aracının ölçmek istediği yapıyı ölçüp ölçmediği ölçeği geliştiren kişilerin kendilerine değil, uzman kararlarına bırakılmıştır.

Kapsam geçerliliği mantıki ve istatistiki olmak üzere iki yolla araştırılabilir.

i) Mantıki yoldan ölçeğin kapsam geçerliliği: Ölçek uygulanmadan ölçeğin geçerliliği tahmin edilir. Ölçüm konusu kavramsal olarak tanımlanır. Ölçekteki her maddenin ve bunların dağılımının ölçüm konusunu örnekleyip örneklemeyeceği araştırılır.

ii) İstatistiki yoldan ölçeğin kapsam geçerliliği: Daha önceden geliştirilen ve geçerliliği yapılmış bir ölçek ile yeni geliştirilen ölçek aynı anda bireylere uygulanır ve bu iki ölçeğin ilişki katsayısı hesaplanır ve bulunan değer ölçeğin geçerlilik katsayısı olarak nitelendirilir.

Uzmanlardan ölçekteki her bir maddenin ölçme derecesini 100 puan üzerinden değerlendirmeleri istenir. Alınan uzman görüşleri doğrultusunda yapılan ifade değişikliklerinin ardından Kendall Uyuşum Katsayısı (Kendall Coefficient of Concordance W) korelasyon testi uygulanarak ölçeğin kapsam geçerliliği çalışması yapılır. $P > 0,05$ uzmanlar arasında bir uyum olduğunu gösterir. (SPSS:Non parametric test>k related samples)

Kapsam Geçerliliğinin Aşamaları :

- i. Kavramsal yapı ve test evreninin tanımlanması
- ii. Konunun kavramsal boyutlarının ortaya çıkarılması
- iii. Maddelerin belirlenen boyutlara göre oluşturulması
- iv. Geliştirilen ölçeğin uzmanlara değerlendirilmesi
- v. İstatistiksel ve matematiksel analizlerin yapılması

Lawshe'nin İçerik Geçerliliği Oranı:

Bu oran formülü ile hakemlerin her bir ifadeyi (maddeyi) nasıl değerlendirdikleri hesaplanır.

$$İGO_i = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

$İGO_i$: Ölçeğin i'nci maddesinin içerik geçerliliği oranı

n_e : İfadenin uygun olduğunu belirten hakem sayısı

N : Toplam hakem sayısı

İGO -1 ile +1 arasında bulunur. Hakemlerin yarısından daha azı bir madde için “uygun” işaretlemesi yapmışsa sonuç – çıkar. Eksi işaretli maddeler ölçekten çıkarılmalıdır. Hakemlerin %50'si bir madde için uygun yada uygun değil demişse İGO sıfır çıkar. Bunun için bir maddenin ölçekte yer alması için hakemlerin %50'den fazlası uygun demelidir. İGO hakemler arasındaki uyuşmanın tesadüfe bağlı olup olmadığı belli olmaz. İGO sadece değerlendiriciler arasındaki uyuşmayı gösterir.

Örnek.

Maddeler	Uygun	Kalabilir	Uygun Değil	Toplam Hakem
1	10	2	0	12
2	2	10	0	12
3	0	2	10	12
4	8	2	2	12
5	8	3	1	12
6	7	4	1	12
7	8	2	2	12
8	8	2	2	12
9	6	3	3	12
10	4	4	4	12

Maddeler	n_e	N/2	İGO	Karar
1	12	6	1,00	Kabul
2	12	6	1,00	Kabul
3	2	6	-0,66	Red
4	10	6	0,66	Kabul
5	11	6	0,83	Kabul
6	11	6	0,83	Kabul
7	10	6	0,66	Kabul
8	10	6	0,66	Kabul
9	9	6	0,50	Kabul
10	8	6	0,33	Kabul

3. Kriter (Referans) Geçerliliđi (Criterion validity):

Bu geçerlilik türü test puanlarının belirlenen bir veya birkaç dış ölçütle ilişkisini inceler.

Ölçek sonuçlarının aynı kavramsal yapıyı ölçen benzer diğer ölçüm sonuçlarıyla (Benzer ölçekler) tutarlı olması istenir. Benzer bir ölçek aynı deneklere uygulanır ve sonuçlar arasındaki korelasyona bakılır.

i) Kestirim (Yordama) Geerlilięi (Predictive Validity)

Bir leęin kestirim geerlilięi, o lekten elde edilen tahmini puan ile llmek istenen zellikleri ltę bilinen kriter arasındaki korelasyonun hesaplanmasıyla elde edilir.

niversiteye ęrenci alırken ęrencinin lisedeki baęarısının veya niversite giriř sınavındaki baęarısının niversite eęitimindeki baęarısı ile paralel olacaęı dřncesi vardır. niversite giriř testinin niversitedeki performansını tahmin ettięini gsterebilmek iin niversite giriř testinin tahmin geerlilięine sahip olması gerekir. Bu nedenle de Yapılan test ile ilerdeki performans arasında iliřki kurulmak istendięi durumlarda testin tahmin geerlilięinin llmesine gereksinim vardır.

Eęer istatistik dersi iin yapılan bir yılsonu sınavının sonuları banka giriř sınavında sorulan genel istatistik soruların sonuları ile uyumlu deęilse kestirim geerlilięi zayıftır denir.

i) Kestirim (Yordama) Geerlilięi (Predictive Validity)

Kestirim geerlilięi standart test kullanılarak belirlenir. Kestirim geerlilięi standart lümü kestiren test ya da deęişkenlerden elde edilen bir skoru kullanarak standardın kestirilme derecesidir. rneęin LYS sınavından elde edilen puanlar adayların gelecekteki başarısının kestiricisi olarak kullanılır.

İyi bir kestirim deęerine ulaşmak için kestirici (baęımsız- x) deęişken (ler) ile referans (kestirilen- y) deęişken arasındaki korelasyon katsayısının yüksek olması istenir. Ayrıca regresyon denklemi yardımıyla belli bir x için y deęeri kestirilebilir. Burada temel sorun bulunan kestirimin ne kadar güvenilir olduğudur. Bunun için rneklem büyüklüęü ve regresyon denkleminin standart hatasından yararlanılabilir.

Bu tür alışmalarda bir dięer sorun elde kestirim denkleminin yeni bir veri grubuna (rnekleme) uygulanması durumunda, geerlilik katsayısının (korelasyon) azalma eęiliminde olup olmadığıdır. Bunun için [apraz geerlilik \(cross validation\)](#) yönteminden yararlanılır.

ii) Uyum (Yakınsama) Geçerliliği (Concurrent Validity)

Aynı yapıyı ölçmek için kullanılan alternatif araçların sonuçlarının birbirlerine yakınsaması veya aynılığının bir ölçüsü olarak kullanılan bir geçerlilik şeklidir.

Ölçmek istediği özelliği kesin bir şekilde ölçen ve yüksek geçerliliğe sahip testlere altın standart ya da referans test denir. Geliştirilen test ile referans test arasındaki korelasyon yüksekse uyum geçerliliği yüksektir denilir.

Yeni geliştirilen bir teste ilişkin uyum geçerliliği için aşağıdaki süreç izlenir.

- Altın Standart-referans test tanımlanır.
- Veriler nicel ise referans test ile yeni geliştirilen test arasındaki korelasyon katsayısı 0,8'den büyük ise yeni testin kullanılabilmesine karar verilir.
- Veriler nitel ise referans teste göre elde edilecek duyarlılık-seçicilik istatistiklerine bakılır.

Nitel Veri İin Geerlilik lütleri

Eđer veriler nitel ise (var-yok, başarılı-başarısız vb.) uyum geerlilięi için Duyarlılık (sensitivity)-Özgüllük (Seicilik-specificity) istatistiklerinden yararlanır.

	Gerek Durum		Toplam
Geliştirilen Test	Başarılı	Başarısız	
Başarılı	A (GP)	B (YP)	A+B
Başarısız	C (YN)	D (GN)	C+D
Toplam	A+C	B+D	N

A (GP): Gerek pozitif; gerekte başarılı olup test sonucunda da başarılı olanların sayısı

B (YP): Yanlış pozitif; gerekte başarısız olup test sonucunda başarılı olanların sayısı

C (YN): Yanlış negatif; gerekte başarılı olup test sonucunda başarısız olanların sayısı

D (GN): Gerek negatif; gerekte başarısız olup test sonucunda da başarısız olanların sayısı

Duyarlılık (sensitivity) Oranı: Gerçek başarılıları ayırabilme yeteneği

$$D=A / (A+C) = GP / (GP+YN)$$

Duyarlılık; gerçekte başarılı olanların % kaçının geliştirilen test sonucunda başarılı olduğunu gösterir ve gerçek başarılar içinde geliştirilen testin başarılıları ayırtedebilme yeteneği olarak tanımlanır.

Özgüllük (specificity) Oranı: Gerçek sağlamları ayırabilme yeteneği

$$Ö=D / (B + D) = GN / (GN + YP)$$

Gerçekte başarısız olanların % kaçının geliştirilen test sonucunda başarısız olduğunu gösterir ve gerçek başarısızlar içinde, geliştirilen testin başarısızları ayırtedebilme yeteneği olarak tanımlanır.

Doğruluk Oranı (accuracy): Testin toplam doğru tanı koyma oranı

$$DO=(A + D) / N = (GP + GN) / (GP +GN + YP + YN)$$

Doğruluk oranı geçerlilik katsayısı demektir. Bu değer 0,5'in altında ise geliştirilen test ile yapılan sınıflandırmanın şans eseri ortaya çıktığı yorumu yapılır. Bu değer 1'e yakın olması istenir.

Pozitif Tanımlama Oranı (predictive value+): Test pozitifler içerisinde doğruluk oranı

$$PV^+ = A / (A + B) = GP / (GP + YP)$$

Test sonucunda başarılı olanların gerçekten başarılı olma olasılığını verir. Pozitif kestirim değeri diye adlandırılır.

Negatif Tanımlama Oranı (predictive value-): Test negatifler içerisinde doğruluk oranı

$$PV^- = D / (C + D) = GN / (GN + YN)$$

Test sonucunda başarısız olanların gerçekten başarısız olma olasılığını verir ve negatif kestirim değeri adını alır.

4.Yapı (Kavram) Geçerliliği (Construct validity):

Yapı geçerliliği doğrudan ölçülemeyen bir özelliği ölçen bir testin ölçme derecesi olarak tanımlanabilir. İlgili yapıyı (psikolojik özellikler doğrudan ölçülemeyen özelliklerdir.) ölçeceği düşünülen değişkenlerin oluşturduğu bütünün, öngörülen yapıyı belirleyip belirlemediğinin incelenmesi gerekir (Alpar, 2016).

Ölçme aracını oluşturan maddelerin kuramsal yapısına ilişkin bir geçerlilik kavramıdır. Bir testteki sorular birbiriyle ilişkili olmalı ve aynı yapıyı oluşturmalıdır. Maddelerin neden-sonuç ilişkileri bir bütünlük içerisinde testinin sonucunu gösterir. Bu geçerlilik türü faktör analizi diye bilinen istatistik yöntemle analiz edilerek bulunur. Bu analizde amaç ilişkili maddeleri bir araya toplayarak yeni sanal maddeler (faktörler) oluşturmaktır. Yada önceden geçerliliği saptanmış bir araç ile yeni ölçme aracı karşılaştırılarak veya sonucu bilinen bir gruba uygulanarak yapı geçerliliği belirlenebilir.

4.Yapı (Kavram) Geçerliliği (Construct validity):

Ölçme aracı iki gruba uygulanır, bu gruplardan birisi sonuçları bilinen bir gruptur. Ölçme sonunda bilinen grubun yeni ölçme aracı ile beklenen sonuçları uyumlu ise ölçme aracının yapısal geçerliliği olduğu söylenebilir. Genelde bu şekilde bir grubu bulmak zor olduğundan yapısal geçerlilik faktör analizi ile karara bağlanır.

Aslında yapı geçerliliği araştırmadaki her kavramın ölçmeye başlamadan önce çok net tanımlarının yapılması ile yakından ilişkilidir.

Testin ölçülmek istenen davranış bağlamında soyut bir kavramı (faktörü) doğru biçimde ölçebilme derecesidir. Bireyin tutum, performans, yetenek gibi özelliklerini ölçmek için sorular sorulabilir. Hazırlanan bu soruların belirtilen özellikleri ne ölçüde doğru ölçtüğü yapı geçerliliği ile araştırılır.

Yapı geçerliği bir testin içerdiği boyutların ve maddelerin belli bir teorik yapıyı (kavramı) ölçmeye yönelik olduklarının ortaya konulması ile ilgilidir (Anastasi, 1982; Öner, 1994). Yapı geçerliği, gerecin değerlendirmesi için düzenlendiği olguyu değerlendirilme derecesinin ölçer (Creswell, 1994). Yapıyla bağlantılı kategoride sınıflanan bilgiler, test puanlarını, ilgilenen psikolojik özelliğin ölçümü olarak ele alır. Bir testin ilgilendiği yapı, kavramsal bir çerçeve içinde ele alınmalıdır. Kavramsal çerçeve yapının anlamını, diğer yapılardan farklarını belirler ve bu yapıya ait ölçümlerin diğer yapılarla nasıl bağlantılı olması gerektiğini gösterir (Hovardalıoğlu, Sezgin, 1998).