

GEÇERLİLİK VE GÜVENİLİRLİK

Prof. Dr. Bektaş TEPE

Bu bölümde;

- Ölçme sürecinin başarılı bir biçimde yürütülebilmesi için veri toplama araçlarının geçerli ve güvenilir olması gerekmektedir.
- Bu bölümde güvenilirlik ve geçerlilik kavramları, ilişkili oldukları diğer kavramlarla birlikte ele alınacaktır..

Ölçme nedir?

- Herhangi bir objenin belirli bir özelliğini, belirli kurallara dayalı olarak, sayarak, sınıflandırarak, derecelendirerek ya da birimlerle sayısal olarak ifade etme sürecidir.

Ölçmenin önemi

- Hemen hemen bütün bilim dalları deneysel ve kuramsal olmak üzere iki yöne sahiptir ve bu iki yönün ilişkisi ölçme ile sağlanır.
- Ölçme araçları kullanılarak sembolleştirilen kavramlar genelleştirilebilir ve bilimdeki gelişmelere katkıda bulunarak bilime yön verebilir.
- Bir ölçme aracının standart olarak kabul edilebilmesi için sahip olması gereken özellikler vardır.
- Ölçme aracını standartlaştıran ön önemli özelliklerden ikisi güvenilirlik ve geçerliliğdir.

Güvenilirlik

- Veri toplama aracının farklı zamanlarda ölçme yaptığında, ölçtüğü özelliğe ilişkin benzer değerler vermesidir.

Ölçme aracının güvenilir olup olmadığını anlamak

- Ölçme aracının güvenilirlik özelliğine sahip olup olmadığını anlamak için aynı bireylere uygulanan iki ölçümden elde edilen veriler arasındaki korelasyona bakılır.
- Bulunan korelasyon katsayısı güvenilirlik katsayısı olarak isimlendirilir.
- Güvenilirlik katsayısının 0.70'ten büyük olması beklenir.
- Bu katsayı arttıkça ölçme aracının hata oranı da azalmaktadır.

Ölçmenin amacı nedir?

- Ölçme sonuçları tamamen hatalı olamayacağı gibi, tamamen hatasız da olamaz.
- Ölçmedeki hatalar ölçme aracından kaynaklanabileceği gibi, ölçme aracı dışındaki faktörlerden de kaynaklanabilir.
- Bu durumda, her ölçümde bir miktar hata bulunabileceği söylenebilir.
- Ölçmede asıl hedef hatasız sonuçlar elde etmek değil, hataları en aza indirmektir.
- Hata puanları azaldıkça gözlenen değer gerçek değere yaklaşır ve ölçüğün güvenilirliği artar.

Hata eřitleri

- Sabit hatalar
- Seri hatalar
- Rastlantısal hatalar

Sabit hatalar

- Her ölçümde aynı miktarda ve aynı yönde görülen hatalardır.
- Bu tür hataların kaynağı ölçeğin kendi yapısındadır.
- Örneğin; bir tartının her defasında 500 gr eksik ya da fazla tartması sabit hatadır.

Seri hatalar

- Belirli bir yöndedir ve ilerleyici ya da gerileyici özellikte olan hatalardır.
- Bireylerdeki gerçek dışı eğilimlerden kaynaklanır.
- Örneğin; yaşlanan insanın unutkanlığının artması, puanlarda ilerleyici bir şekilde düşmeye sebep olur.

Rastlantısal hatalar

- Yönü ve miktarı bilinmeyen hatalardır.
- Değişen kişisel koşullardan, çevresel koşullardan veya ölçüm aracındaki değişimlerden kaynaklanabilir.

Güvenilirlik testini sınırlayan faktörler

- Ölçme araçlarının güvenilirliklerini sınamak için geliştirilen bazı yöntemler vardır.
- Ancak bu yöntemleri sınırlayan bazı faktörler bulunmaktadır.
- Testin niteliği, zaman, maddi kaynaklar bu faktörler arasındadır.
- Testin içeriğinin karmaşık olması, yeterli zaman olmaması, daha fazla masrafa sebep olması gibi faktörler de birden fazla yöntemin uygulanmasını zorlaştırmaktadır.

Güvenilirliđi kanıtlama yöntemleri

- Test-tekrar test yöntemi
- Puanlayıcılar arası güvenlik
- İç tutarlılık
- Paralel (eřdeđer) formlar yöntemi

Test-tekrar test yöntemi

- Bu yöntemde, veri toplama aracının kararlılığını belirlemek için öncelikle pilot çalışma için bir örneklem seçilir.
- Katılımcılara veri toplama aracı uygulanır ve her birey için puanlar hesaplanır.
- Aradan biraz zaman geçmesine izin verilir.
- Ne kadar zaman geçmesi gerektiğine ilişkin ölçüt, katılımcıların soruları hatırlayamayacakları kadar sürenin geçmesidir.
- Aynı aracın ikinci kez uygulanmasının ardından her birey için puanlar karşılaştırılır.

Test-tekrar test yöntemi

- İki sonuç arasındaki korelasyon hesaplanır.
- Hesaplanan korelasyon katsayısı ölçme aracının güvenilirlik katsayısıdır.
- Hesaplanan korelasyon katsayıları -1 ile +1 arasında değerler alır.
- Korelasyon +1 ise pozitif -1 ise negatif ilişki bulunduğu ifade edilir.
- 0 ise herhangi bir korelasyonun bulunmadığını gösterir.

Puanlayıcılar arası güvenilirlik

- Bu yöntemde, birden fazla puanlayıcı/hakem birbirinden bağımsız olarak aynı durumu aynı ölçme aracıyla puanlar.
- Puanlama sonunda, puanlayıcıların verdiği puanların tutarlılığına yani birbiriyle olan uyumuna bakılır.
- Verilen puanlar arasındaki korelasyon hesaplanır.
- Korelasyon katsayısının yüksek ve pozitif yönlü olması beklenir.

Puanlayıcılar arası güvenilirlik

- Bu durum, verilen puanların tutarlı olduğunu ve ölçme aracının güvenilir olduğunu gösterir.
- Verileri anlamlandırmak için korelasyon katsayısının yanı sıra t testi, Kendall'ın Konkordasyon sayısı ve Cronbach Alfa kullanılır.
- Puanların aritmetik ortalama ve standart sapmalarının birbirine benzer olması da güvenilirlik derecesinin yüksek olduğunu gösterir.

İç tutarlılık

- Testin bir defa uygulanmasına dayalı olarak test maddelerine verilen cevaplar arasındaki ilişkileri belirlemektir.
- Ölçek, bir gruba bir kez uygulandıktan sonra maddeler yansız olarak veya '*çift numaralılar-tek numaralılar*' gibi bir yöntemle iki yarıya bölünür.
- İki yarıdan elde edilen puanlar uygun bir yöntemle değerlendirilir.

İç tutarlılık

- Eğer şartlar sağlanıyorsa Pearson Momentler Çarpımı Korelasyon katsayısı hesaplanır.
- Ölçeğin bütünüyle ilgili güvenilirlik katsayısını hesaplamak için ise Spearman-Brown formülü yerine Cronbach Alfa Katsayısı kullanılır.

Paralel (eşdeğer) formlar yöntemi

- Birbirine alternatif formların aynı gruba, aynı koşullarda, aynı ya da farklı zamanlarda uygulanmasıdır.
- İki formun eşdeğerliliğinin (formlardaki madde sayısı, niteliği, kullanılan ölçekleme tekniği, faktör yapısı, ölçülen davranış vb.) denk olması gerekir.
- Her iki araçtan elde edilen puanlar bütün grup için ayrı ayrı hesaplanarak aralarındaki ilişki belirlenir.
- Hesaplanan korelasyon katsayısının yüksek ve pozitif yönlü olması beklenir.
- Bu durum, ilişkinin yüksek derecede olduğunu gösterir.

Paralel (eşdeğer) formlar yönteminin zorlukları

- İki formu eşdeğer hale getirmek güç olabilir.
- Bu işlem sırasında araştırmacının daha fazla zamana ihtiyacı olabilir.
- Uygulama daha çok çaba gerektirir.
- Bunlara paralel olarak masraflar da artabilir.

Geçerlilik

- Bir ölçme aracının, ölçmeyi amaçladığı özelliği, başka herhangi bir özellikle karıştırmadan ölçebilmesi onun geçerliliğini gösterir.
- Geçerlilik katsayısı -1 ve $+1$ arasındadır.
- Eğer ölçme aracı hedeflenen özelliği iyi bir şekilde ölçerse, yapılan çıkarımların ve yorumların isabetli olma olasılığı artar.

Geçerlilik derecesini etkileyen faktörler

- Bir ölçeğin geçerlilik derecesi
 - Kullanılış amacına
 - Uygulandığı gruba
 - Uygulanma ve puanlama biçimine göre farklılık gösterir.
- Ölçülmek istenen özelliğin iyi tanımlanması, yani doğru ifade edilmesi de geçerliliğin derecesini etkiler.

Geçerlilik kategorileri

- Kapsam geçerliliği
- Ölçütle alakalı geçerlilik
 - Yordama geçerliliği
 - Eşzaman geçerliliği
- Yapı geçerliliği

Kapsam geçerliliği

- Ölçeğin, ölçmek istediği özellik dışındaki faktörlerden arınmış olması anlamına gelmektedir.
- Amaç, ölçeğin bütününe ve alt boyutlarının ölçülmek istenen alanı ölçüp ölçmediğini ve ölçülecek alan dışında farklı kavramları barındırıp barındırmadığının değerlendirilmesidir.
- Örneğin; öğrencilerin yabancı dilde dilbilgisi yeterliliğini ölçmek isteyen bir testin, yabancı dilde kelime yeterliliğini ölçen bir madde bulundurmaması gerekir.

Ölçütle alakalı geçerlilik

- Bir ölçme aracının diğer bir ölçme aracıyla elde edilen sonuçları verme yeteneği olarak tanımlanır.
- Aynı zamanda, ölçme aracının, ölçmeyi amaçladığı özelliği ne kadar başarıyla tahmin ettiği sorusuna da yanıt verir.
- Uygun korelasyon yöntemiyle elde edilen ilişkinin derecesi '*geçerlilik katsayısı*' olarak adlandırılır.

a) Yordama (tahmin) geçerliliği

- Eğer bir ölçek, puan hakkında tahminde bulunmak, ileride ne olacağına karar vermek gibi bireylerin ilerideki performansını tahmin etmek için kullanılacaksa, ölçeğin yordama geçerliliğine bakmak gerekmektedir.
- Ölçme aracı, ilgili gruba uygulandıktan sonra, ölçülen davranışın belirgin olarak ortaya çıkmasına yetecek kadar belirli bir süre beklenir.
- Bu süre geçtikten sonra, ölçüt puanları ile ölçekten elde edilen puanlar arasındaki ilişki katsayısı hesaplanır.
- Hesaplanan katsayı ne kadar yüksekse, ölçme aracının geleceğe yönelik tahmin gücü de o kadar yüksek olacaktır.

b) Eşzaman geçerliliği

- Bu yöntemde, bir gruba geçerliliği sınanmak istenen yeni bir ölçek ve önceden geçerliliği saptanmış eski bir ölçek aynı zamanda uygulanır.
- Uygulama sonunda, bireylerin, güvenilirliği saptanmış olan ölçekten aldığı puanlarla, yeni geliştirilen ölçekten alınan puanlar arasındaki korelasyon katsayısı hesaplanır.
- Katsayının yüksek olması beklenir.
- Bu durum, yeni geliştirilen ölçeğin geçerliliğinin yüksek olduğunu gösterir.

Yapı geerlilięi

- Kullandığımız araların yapıları (örn; zeka, cinsiyet, yař, depresyon, öz-yeterlilik, kiřilik, yeme bozukluęu, patoloji, biliřsel stiller vb.) ne ölçüde temsil ettięi konusunda çıkarımda bulunma işlemidir.

Yapı geçerliliği

- Örneğin; fakir bölgelerde yaşayan insanların evliliklerinde yaşadıkları fikir ayrılıklarında depresyonun etkileri araştırıldığında,
 - katılımcılar (dezavantajlı insanlar),
 - bağımsız değişken (depresyon),
 - bağımlı değişken (evlilikteki fikir ayrılıkları) ve
 - araştırma ortamı (fakir mahalle)

gibi yapıların incelenmesi gerekmektedir.

- Bu şekilde adım adım elde edilen bilgilerden yola çıkılarak ölçeğin yapısı hakkında bilgi elde edilir.

Geçerlilik ve güvenilirlik ilişkisi

- **Güvenilirlik** ölçme aracının kararlılığı, tutarlılığı, değişkenin boyutunu hatalardan ne kadar arındırarak ölçtüğü ile ilgilidir.
- **Geçerlilik** ise, ölçme aracının diğer değişkenleri karıştırmadan ölçmeyi amaçladığı değişkeni ölçüp ölçmediğini, yani doğru işi yapıp yapmadığını sınar.

Geçerlilik ve güvenilirlik ilişkisi

- Bir ölçme aracının geçerlilik özelliğine sahip olabilmesi için güvenilirlik özelliğine sahip olması şarttır.
- Fakat sadece güvenilirliği sağlaması, o ölçme aracını geçerli yapmaz.
- Bilimsel araştırmalarda kullanılan ölçme araçlarının bu iki özelliği taşıması çok önemlidir.
- Bu özelliklere sahip olmayan testleri kullanmak, araştırma sonunda elde edilen verilerin ve bulguların geçerli kabul edilmesini ya da herhangi bir amaçla kullanılmasını imkansız kılacaktır.

SABRINIZ İÇİN TEŐEKKÜR EDERİM

Prof. Dr. Bektař TEPE