

## GÜVENİLİRLİK ANALİZİ

- Güvenilirlik tutarlılık demektir. Aynı bireylerin farklı zamanlarda aynı sonuçları vermesi durumudur.
- Güvenilirlik ölçütleri genellikle 0 ile 1 aralığında değer alır.
- Güvenilir ölçütlere örnek olarak kişilere ait boy ve ağırlıklar verilebilir.
- Güvenilirlik, geçerlilik anlamına gelmez. Bir şeyin tutarlı olması, gerçekten ölçülmek istenilen şeyin ölçüldüğüne işaret etmez. Fakat güvenilir olmayan bir test geçerli olamaz.

### Güvenilirliği kontrol etme yöntemleri:

#### Değerlendirici arası güvenilirlik (Inter-rater reliability)

Değerlendirici güvenilirliği, iki veya daha fazla değerlendirici arasındaki eşleşme derecesini ölçer. Bu yönteme göre örneğin mide ağrısı çeken bir kişi farklı doktorlara gittiğinde tüm doktorların aynı teşhisi koyması durumu, bu teşhisi güvenilir kılar. Bu yöntemde üç farklı güvenilirlik şekli mevcuttur:

1. Değerlendiriciler, resmi performans puanları ile hemfikir olduklarında güvenilirdir.
2. Değerlendiriciler, verilecek kesin puanlar konusunda birbirleriyle hemfikir ise güvenilirdir.
3. Değerlendiriciler, hangi performansın daha iyi hangisinin daha kötü olduğu konusunda hemfikir ise güvenilirdir.

Değerlendiricilerin davranış tanımlarına göre bu güvenilirlik ölçütleri ikiye ayrılmaktadır. Bu davranış tanımları şöyledir:

1. Güvenilir değerlendiriciler, birer “oylama makineleri”dir. Örnek olarak yazılı kağıtlarını okuyan bilgisayarlar verilebilir. Bu davranış genelleştirilebilirlik teorisi ile değerlendirilebilir.
2. Güvenilir değerlendiriciler, bağımsız gözlemcilerdir. Bağımsızlıklarını hafif anlaşmazlıklarla gösterirler. Bu davranış, Rasch modeli ile değerlendirilebilir.

Bu yöntemde kullanılan ölçütlerin bazıları Ortak Anlaşma Olasılığı, Kappa İstatistiği, Korelasyon Katsayıları, Sınıflar İçi Korelasyon Katsayısı, Anlaşma Sınırları Ve Krippendorff's Alpha ölçütleridir.

#### Test-Tekrar test güvenilirliği (Test-retest reliability)

Test-Tekrar test güvenilirliğinde tek bir değerlendirici aynı yöntemi veya aracı aynı koşullarda farklı zamanlarda iki kez uygular. Bu şekilde her kişi için iki farklı puan elde edilir ve bu iki puan kümesi arasındaki korelasyon, test-tekrar test güvenilirlik katsayısını verir. Test güvenilirse, puanlar arasında yüksek pozitif bir ilişki olacaktır.

Tekrarlanabilirliğin sağlanması için aşağıdaki koşulların sağlanması gerekmektedir:

- Aynı deneysel araçlar
- Aynı gözlemci
- Aynı ölçüm aleti, aynı koşullarda
- Aynı yer
- Tekrar kısa bir süre içerisinde yapılmalı

- Aynı amaç

### Yöntemler arası güvenilirlik (Inter-method reliability)

Yöntemler arası güvenilirlik, kullanılan ölçüm aletleri veya yöntemler arasında değişim varken ne kadar tutarlılık mevcut olduğunu ölçmede kullanılır. Formlarla yapılırken Paralel Formlar güvenilirliği olarak adlandırılır. Bu yöntem, aynı grupta iki farklı formdan alınan puanlar arasındaki ilişkiye bakılarak uygulanır.

### İç tutarlılık güvenilirliği (Internal consistency reliability)

İç Tutarlılık Güvenilirliğinde bir test içerisindeki birimler boyunca mevcut olan tutarlılık hesaplanır. Aynı genel yapıyı ölçtüğü düşünülen bir çok birimin benzer skorlara sahip olup olmadığını ölçer. Örnek olarak bir yanıtlayıcı “bisiklet sürmeyi severim” ve “geçmişte bisiklet sürerken keyif aldım” sorularına katılıp “bisiklet sürmekten nefret ederim” sorusuna katılmıyorsa, bu tutarlılığa işaretir.

İç Tutarlılık güvenilirliği Cronbach’s alpha ölçütü ile ölçülür. Bu istatistik birimler arasındaki ikili korelasyonlar üzerinden hesaplanır. Cronbach’s alpha şu şekilde hesaplanır:

$$\alpha = \frac{K}{K-1} \left( 1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_X^2} \right)$$

Burada  $K$  madde sayısını,  $\sigma_X^2$  toplam test skorunun varyansını ve  $\sigma_{Y_i}^2$  ise  $i$ -nci maddenin toplam örneklemdaki varyansını gösterir. İç tutarlılık 0 ile 1 aralığında değişir. alpha katsayısı, birimler içi varyans birimler arası varyansı geçtiğinde negatif değer alabilir. Cronbach’s alpha değeri genellikle aşağıdaki gibi yorumlanır.

Cronbach’s alpha	İç tutarlılık
$0,9 \leq \alpha$	Harika
$0,8 \leq \alpha < 0,9$	İyi
$0,7 \leq \alpha < 0,8$	Kabul edilebilir
$0,6 \leq \alpha < 0,7$	Şüpheli
$0,5 \leq \alpha < 0,6$	Zayıf
$\alpha < 0,5$	Kabul edilemez

Çok yüksek güvenilirlikle (Örneğin 0,95 ve yukarısı) çok istenilen bir durum değildir. Çünkü bu durum gereksiz birimlerin olduğuna işaret eder. Burada amaç benzer skorlar elde etmeni yanı sıra her bir birimin kendine has bir katkısının olmasıdır.