

6. ARAŞTIRMALARDA ÖLÇME VE ÖLÇEKLER

Birlikte Düşünelim

- Bilimsel araştırmalarda geçerlilik ve güvenilirlik ne anlama gelir? Bunlar niçin önemlidir?
- Araştırmalarda ölçmenin gerekliliği nedir?
- Ölçme sırasında ne gibi hatalarla karşılaşılabilir?

Başlamadan Önce

Dördüncü bölümde, bilimsel araştırmalarda veri konusu ele alındıktan sonra birincil ve ikincil veri kaynakları ile anket, mülakat ve gözlem konuları ele alınmıştı. Bu bölümde ise, araştırmalarda en önemli konulardan olan ölçme ve ölçekler üzerinde durulacaktır. Çeşitli ölçek türleri incelendikten sonra ölçeklerle ilgili önemli hususlara değinilecektir.

Giriş

Bilimsel araştırmalarda bilimselliğin temel şartı araştırılan kavramlar, konular veya değişkenlerin ölçülebilir nitelikte olmasıdır. Dolayısıyla değişkenlerin, kavramların veya olguların ölçülmesi, bilimsel araştırma sürecinin ayrılmaz bir parçasını oluşturur. Ölçülmesi gereken değişken fiziksel bir yapı olduğunda (ağırlık, mesafe, yoğunluk vb.) iş görece olarak kolaydır. Yine deneğin yaşı, cinsiyeti, eğitim ve gelir düzeyi gibi değişkenlerin rakamsal ifadesi de mümkündür.

Ancak, tutumlar, değerler ve algılar gibi sosyal bilimlerde incelemeye konu olan kavramlara/yapılara/değişkenlere kesin değerler atfetmek kolay değildir. Çünkü bu değişkenlere değer atfetme cevaplayıcının sübjektif değerlendirmesine bağlı olduğundan, yapılacak tüm değerlendirmeler mutlak objektiflikten uzak olacaktır. Sözelimi ölçülmesi gereken 'motivasyon ile performans arasındaki ilişki' ise bu durumda performans ölçütü olarak üretilen parça/saat gibi bir değer oluşturulabilir ancak motivasyon ölçütü olarak hangi değerler ele alınabilir? Dolayısıyla bazı değişkenleri dolaylı yollardan göstergeleri veya yansımaları üzerinden ölçmek zorundayız.

6.1. Ölçme ve Ölçüm Düzeyleri

Basit bir ifadeyle ölçme, *önceden tespit edilen belirli kurallara göre nesnelere ve kişilerin özelliklerine sayılar veya semboller atamaktır*. Buna karşılık ölçekler, deneklerin belli veriler baz alınarak sınıflandırıldıkları araç veya mekanizmalardır.

Ölçüm Düzeyleri, Tipik Örnekler ve Uygulanabilecek Temel İstatistikler

Ölçüm Düzeyi	Temel Mukayese	Tipik Örnekler	Ortalama Değer
Nominal/Sözde	Kimlik	✓Cinsiyet ✓Göz/saç rengi ✓Meslekler ✓Araba plakaları ✓Sporcu forma numaraları	Mod
Ordinal/Sıralı	Sıra/Sıralama	✓Marka tercihi ✓Toplumsal sınıf ✓Bölüm tercihi ✓500 büyük firma	Medyan
Interval/Aralı	Aralıklar	✓Sıcaklık ölçeği ✓Başarı puanı ✓Markaya karşı tutum	Aritmetik ortalama
Rasyo/Oranlı	Mutlak büyüklükler	✓Satış miktarı ✓Müşteri sayısı ✓Ağırlık ✓Zaman ✓(mutlak değeri olan her şey)	Bütün matematiksel işlemler

Araştırma yöntemleri bağlamında dört temel ölçüm seviyesi vardır. Bunlar, *sözde (nominal)*, *sıralı*, *aralı* ve *oranlı* ölçekler. Sözde ölçekten oranlı ölçeğe gidildikçe ölçeğin ölçme gücü (verilerin kalite düzeyi) artar. Yani, düşük ölçüm düzeyinde (örneğin nominal ölçek) elde edilen veriler, araştırmacıya pek fazla işlem/analiz yapma imkanı tanımadığından olguyu tanımlama gücü sınırlıdır. Bu yüzden araştırmacı koşullar elverdiğince yüksek düzeyde ölçüm yapmaya çalışmalıdır. Yukarıdaki tabloda değişik ölçüm düzeyleri ve bunlara ait örnekler yer almaktadır.

Nominal Ölçüm: Sayılar ve nesnelere arasında sadece bir ilişkinin öngörüldüğü, nesnelere sadece gruplandırıldığı bir ölçme seviyesidir. Bu seviyede sayılar ve harfler, sadece ve sadece bir kimlik olmaktan öteye geçmemektedir. Önemli/önemsiz, astlık üstlük oluşturulmadığından bir sıralama yapmak da mümkün değildir. Bu seviyede yapılan ölçüm değerleri için genellikle sadece frekans dağılımını yapmak mümkündür. İİBF öğrencilerini bölümlerine göre gruplandırmak, araba plakaları, doğum yeri, cinsiyet, göz rengi vb. hep nominal türden ölçeklerdir.

Sıralı Ölçüm: Bu düzeydeki bir ölçüme bakarak, bir nesnenin belirli bir özelliği az mı yoksa çok mu taşıdığını söylemek mümkündür. Sıralı ölçümlerde nesnelere göre konumları belirlenebilir, fakat nesnelere arasındaki farkın mutlak boyutunu belirlemek mümkün değildir. Örneğin, üç yarışçı (A, B, ve C) yarışı sırasıyla 1., 2., ve 3. olarak tamamlamaktadır. Sıralı ölçüm seviyesi sadece yarışmacıların 1., 2. ve 3. olarak sıralandığını göstermektedir. Yarışçılar arasında (yani 1. ve 2. yarışçı veya 2. ve 3. yarışçı arasında) ne kadar bir süre farkının olduğu konusunda bir şey söylemek imkansızdır. Aynı şey 500 Büyük Firma listesinde ilk üç sırayı sözcümleri (1)TÜPRAŞ, (2)TEAŞ ve (3)TEKEL'in alması örneğinde de görülür. Bu örnekte de şirketler arasında satışlar bazında sıralama yapılmakta ancak, aralarındaki farkın ne olduğu bilinmemektedir. Uygulanabilecek işlemler arasında medyan ve Spearman Rank korelasyonu sayılabilir.

Aralı Ölçüm: Sayısal olarak eşit aralıkların eşit mesafeleri temsil ettiği bir ölçek olup nesnelere sıralanmasında kullanılmaktadır. Bu ölçüm seviyesinde 1 ile 2 arasındaki mesafe ile 6 ile 7 arasındaki mesafenin birbirine eşit olduğu varsayılmaktadır. Ordinal ölçümün bütün özelliklerini taşımaktadır. Ayrıca, bu ölçüm seviyesi nesnelere arası farkların mukayesesine de imkan tanımaktadır. Aralık ölçeğinde sıfır noktasının yeri sabit değildir, yani başlangıç noktası görecelidir. Bu sebeple iki değer oranını almak anlamlı olmayacaktır. Zira buradaki 0 (sıfır) değeri aslında gerçek sıfır değildir. Örneğin, 'oda sıcaklığı 0 derecedir' dendiğinde buradaki sıfır gerçek değildir. Odada sıcaklığın olmadığı söylenemez. Zira 0 derece -10 dereceden 10 derece daha fazla ısı anlamına gelmektedir. Sıfır gerçek olmadığı için aritmetik karşılaştırmalar da yapılamaz. Örneğin, IQ'su 140 olan birisi 70 olandan iki kat daha zekidir denilemez. Hemen hemen bütün istatistiksel testler bu seviye ölçüme uygulanabilmektedir.

Likert ölçeği, semantik farklar ölçeği ve değişen sayıda aralıklardan oluşan görece önem özellikleri. Aşağıda bir Likert ölçeği örneği görülmektedir.

Aşağıdaki ifadelere ne derecede katılıyorsunuz?	Hiç katılmıyorum	Kısmen katılmıyorum	Emin değilim	Kısmen katılıyorum	Tamamen katılıyorum
<i>Yaptığım iş bana yeteneklerimi gösterme şansını veriyor.</i>	1	2	3	4	5
<i>Bu iş benim için çok şey ifade ediyor.</i>	1	2	3	4	5
<i>Farklı bir işte çalışmayı isterdim.</i>	1	2	3	4	5

Oranlı Ölçüm: Daha önceki üç ölçeğin bütün özelliklerini taşımakla beraber, mutlak sıfır noktasının da sahiptir. En üst ölçüm düzeyidir ve her türlü istatistiksel ve matematiksel işleme imkân tanımaktadır. Cevaplayıcı 0'dan herhangi uygun bir sayıya kadar cevap verebilir.

Kaç çocuğunuz var?

Yaşadığınız şehirde kaç süpermarket var?

Yıllık cironuz ne kadar (TL)?

Görüldüğü gibi her düzeydeki ölçek kendisinden düşük düzeydeki ölçekler için kullanılabilir. Kendisinden üst düzeyler için ise kullanılamaz. Örneğin, aralı veriler sıralı düzeyde kullanılabilir ama, oranlı ölçek olarak kullanılmaz. Oranlı veriler ise bütün düzeyler için kullanılabilir. Bu yüzden, veri toplarken mümkün en üst seviyede veriler elde etmek (cevap oranının ve verilerin güvenilirliklerini de dikkate almak kaydıyla) mantıklıdır.

6.2. Ölçek Türleri



Sosyal bilimlerin çeşitli alanlarında çok

değişik türde ölçekler kullanılmaktadır. Aşağıda yaygın olarak kullanılan bazı ölçek tipleri kısaca açıklanarak, bu ölçeğe ait bazı özellikler sıralanmaktadır.

6.2.1. Sürekli Ölçekler

Bu ölçeklerde aşağıda örneklerde görüldüğü gibi, cevaplayıcının, iki uç arasında uzanmakta olan bir çizgi üzerindeki uygun bir yere işaret koymasını beklenmektedir.

Aşağıda üç farklı formatta sürekli ölçeklere ilişkin örnek yer almaktadır. Örneklerde görüldüğü gibi cevaplayıcı iki uç arasında kalan kısma cevabını işaretler. Değerlendirme aşamasında ise araştırmacı cevap çizgisini gereken sayıda aralığa böler ve daha sonra cevaplayıcının cevabını rakamsal olarak ortaya koyar. Bu ölçeklerin oluşturulması kolay olmakla birlikte, skorların belirlenmesi hem zor hem de güvenilir değildir. Pratik uygulamalarda araştırmacı ölçeği oluştururken cevaplama çizgisini 10 cm olacak tarzda ayarlar. Daha sonra araştırmacı deneğin cevabını cetvelle milimetre cinsinden ifade ederek, ölçeğin sürekli ölçek havasına ulaşmasına imkân tanımaktadır.

Diğer marketlerle karşılaştırdığınızda GIMA mağazasını nasıl buluyorsunuz?

Format 1
Muhtemelen en kötüsü X Muhtemelen en iyisi

Format 2
Muhtemelen en kötüsü X Muhtemelen en iyisi
0 50 100

Format 3
Muhtemelen Çok kötü Ne kötü/ne iyi Çok iyi Muhtemelen
En kötüsü 0 10 20 30 40 50 60 70 80 90 100 en iyisi

6.2.2. Tekli Ölçekler

Tek maddeden veya sorudan ibaret olan aşağıda örnekleri görülmekte olan ölçek türüdür. Bunlar arasında *kategorili ölçekler*, *sıralama ölçeği*, *sabit-toplamlar ölçeği*, *şekil/grafik ölçekler* ve *eşli-karşılaştırma ölçekleri* sayılabilir.

6.2.2.1. Kategorili Ölçekler

Bu ölçeklerde cevap seçenekleri kategoriler halinde verilmektedir ve kategoriler arasında çakışma olmamalıdır. Başka bir ifade ile her denek sadece bir seçeneğe ait olabilmelidir. Kategori sayısı ile ilgili bir sınırlama olmamakla birlikte yaygın kullanılan kategori sayıları 2, 3, 5, 7 ve 9'dur. Bazen her kategorinin temsil ettiği anlam açık yazılırken bazen sadece uç kategorilerin yazılmasıyla yetinilmektedir.

Formatına bağlı olarak kategorili ölçeklerden bazıları araştırmacılarca karşılaştırmalı ölçekler olarak da tanımlanmaktadır. Aşağıda bunlara ilişkin örnekler verilmektedir.

Aldığınız nottan ne derece memnunsunuz?
[...] Çok memnun
[...] Öldükçe memnun
[...] Kısmen memnun
[...] Pek memnun değil
[...] Hiç memnun değil

Bu markayı seçmekten memnunsunuz?
Çok memnun 1.....2.....3.....4.....5 Hiç memnun değil

Bölgünüzde bulunan hastanelere kıyasla Sakarya Araştırma Hastanesini nasıl buluyorsunuz?
[...] Çok iyi [...] Vasat/Orta [...] Çok kötü

6.2.2.2. Sıralama Ölçekleri

Bu ölçeklerde, deneklere inceleme konusuyla ilgili bir özellikler listesi verilir ve bu listedeki özelliklerin tercih veya önem sırasına veya başka bir ölçüte göre sıralanması istenir.

Çamaşır makinesi satın alırken, tercihinizin belirlenmesinde aşağıdaki faktörlerin rolünü önem sırasına göre belirtiniz?

Özellikler	Önem sırası
Kalitesi	-----
Fiyatı	-----
Satış sonrası hizmetleri	-----
Ödeme kolaylığı	-----
Markası	-----

6.2.2.3. Sabit-Toplam Ölçekleri

Bu ölçeklerde, deneklere 100 tam puan verilmekte ve bu puanı önem derecesine veya hoşlanma seviyesine göre, (listede) verilen özellikler arasında dağıtması istenmektedir. Bu tür ölçeklerde, özelliklere verilen puanlar toplamı tam olarak 100 olmak durumundadır. Ne aşağı ne de yukarı olabilir.

Bir çamaşır makinesi seçerken göz önüne aldığınız aşağıdaki beş özellik arasında size verilen 100 puanı önem derecesine göre bölüştürünüz.

Kalitesi	-----
Sessiz çalışması	-----
Satış sonrası hizmetler	-----
Ekonomisi	-----
Yedek parça imkânı	-----
TOPLAM	100

6.2.2.4. Grafikselsel/Şekilsel Ölçekler

Bu ölçeklerde cevaplar kelimeler ve harfler ile değil şekiller ve skeçlerle verilmektedir. Özellikle okuma yazma bilmeyen insanlara ve küçük çocuklar üzerinde yapılacak olan anketlerde oldukça yaygın olarak kullanılmaktadır. Aşağıda bunlardan yüz ifadesi biçimine bir örnek verilmiştir.

İşletmemizden aldığınız hizmetlerden ne derece memnunsunuz?



6.2.2.5. İkili-karşılaştırmalı Ölçekler

Bu ölçeklerde deneklerin ikililer arasında seçim yapması beklenmektedir.

Bir ürün satın alırken aşağıdaki özelliklerden hangisi daha önemlidir?

- [...] Kalite - Fiyatı
- [...] Fiyatı - Satış sonrası hizmetler
- [...] Yedek parça - Dayanıklılık
-

6.2.3. Çoklu Ölçekler

Birden çok maddeden meydana gelen ve maddelerin aynı ölçeğe göre değerlendirildiği ölçeklere denir. Daha çok tek boyutlu olarak ölçülmesi mümkün olmayan, çok boyutlu kavramların ölçülmesinde kullanılmaktadır. Bunlar arasında yaygın olarak kullanılanlar *Likert ölçeği* ve *semantik farklılıklar ölçeği* sayılabilir. Bunlar aşağıda kısaca açıklanmaktadır.

6.2.3.1. Likert Ölçeği

Bu ölçek geliştiricisi Rensis Likert'in adıyla anılmakta olup, sosyal bilimler alanında yaygın olarak kullanılmaktadır. Bu ölçekte deneye çeşitli ifadeler ve yargılar yöneltilir. Denekten, bu yargılara veya

ifade-lere katılıp katılmama derecesini belirtmesi istenir.¹ Kategori sayısı olarak 5, 7, 9 ve 11'li olabilen bu ölçek, yaygın olarak 5'li Likert ölçeği tercih edilmektedir. Likert ölçeğinde cevaplar katılıp katılmama ile ilgilidir. Bu ölçek ile aralık seviyesinde bir ölçümün yapıldığı gözlen-mektedir. Unutulmaması gereken son derece önemli nokta her 5'li ölçek Likert ölçeği değildir. Likert ölçeğinin seçenekleri Katılıyorum-Katılmıyorum şeklindeki cevaplardan ibarettir. Likert ölçeği format olarak değişik şekillerde uygulanabilir. Aşağıdakine benzer şekilde tablo şeklinde olabileceği, tek bir soru şeklinde de kullanılmaktadır.

Aşağıdaki ifadelere ne derece katılıyorsunuz?	Hiç katılmıyorum	Kısmen katılmıyorum	Emin değilim	Kısmen katılıyorum	Tamamen katılıyorum
Bu okulu seçmem son derece iyi bir karardı.	1	2	3	4	5
Okulun imkanları son derece sınırlıdır.	1	2	3	4	5
Sosyal yaşantı son derece canlıdır.	1	2	3	4	5
İyi ki bu okulu seçmişim.	1	2	3	4	5

"Bu oteli seçmekle son derece isabetli bir karar vermişim" ifadesine ne derecede katılıyorsunuz?

- [...] Hiç katılmıyorum
- [...] Kısmen katılmıyorum
- [...] Ne katılıyorum ne de katılmıyorum
- [...] Kısmen katılıyorum
- [...] Tamamen katılıyorum

Likert ölçeği daha çok kişilerin tutumlarını ve eğilimlerini ölçme amaçlı olarak kullanılmaktadır. Oluşturma, doldurma ve uygulama açısından avantajlı olan bu ölçeğin en önemli dezavantajı ise doldurma için gerekli olan sürenin görece olarak uzun olmasıdır.

Likert ölçeği ile elde edilen verilerin değerlendirilmesinde iki alternatif bulunmaktadır. Birincisi, ölçekte yer alan her bir ifade ayrı ayrı değerlendirilebilir. İfadeler arası korelasyon veya ilişki incelenebilir. İkincisi, ölçekte yer alan tüm ifadeler için skorların toplanması neticesinde elde edilen kümülatif toplam skor üzerinden değerlendirme yapılmasıdır. Bu sebeple Likert ölçeğine *toplamalı ölçek (summated scale)* olarak da anılmaktadır.

6.2.3.2. Semantik Farklılıklar Ölçeği

Özellikle mamul, marka ve firma imaj çalışmalarında yaygın olarak kullanılan bu ölçekte deneklere, incelenmek istenen nesne, marka, ürüne ait çeşitli özelliklere ilişkin nitelendirmeler zıt kutuplara gelecek şekilde yedi-noktalı bir doğru üzerine yerleştirilmiştir.² Deneklerden her bir özelliğe ait doğru üzerinde kendi değerlendirmelerine uygun bir yere işaret koyarak seçmeleri istenmektedir.

Bu ölçek ile aralık seviyesinde ölçüm yapılmak amaçlanmıştır. Semantik farklar ölçeği 1-7 şeklinde ölçeklendirilebileceği gibi -3- +3 şeklin-de de ölçeklendirilebilir. Elde edilen veriler profil analizi şeklinde yapılır. Her bir niteliğe ilişkin verilen cevaplar ortalama ve medyana bağlı olarak yorumlanmaya çalışılır. Tipik bir semantik farklar ölçeği aşağıda görülmektedir.

Sizce GIMA,

Ucuz !.....! X !.....! Pahalı

Az çeşit !.....! X !.....! Bol çeşit

Geleneksel !.....! X !.....! Modern

6.2.3.3. Stapel Ölçeği

Deneklerin nesnelerin belirli özelliklerine ait değerlendirmeyi aşağıdakine benzer bir ölçek üzerinde işaretledikleri türdür.

GİMA marketleri...		
+5	+5	+5
+4	+4	+4
+3	+3	+3
+2	+2	+2
+1	+1	+1
Ürün çeşitliliği	Fiyat uygunluğu	İç mekan
-1	-1	-1
-2	-2	-2
-3	-3	-3
-4	-4	-4
-5	-5	-5

Görüldüğü gibi, bir araştırmacı ölçmek istediği şeyi (değişkeni) değişik ölçekler kullanarak ölçebilir. Burada bu ölçeklerden yaygın olarak kullanılanları açıklanmaya çalışılmıştır. Ölçeklerin oluşturulması aynı zamanda elde edilecek verinin kalitesini ve düzeyini de belirleyecektir. Sözde ölçekten oranlı ölçeğe gidildikçe elde edilen verinin ve araştırmacının bu ölçeklere uygulayabileceği istatistiksel işlemlerin kalitesi artar.

6.3. Ölçeklerle İlgili Önemli Hususlar

Ölçekler yardımıyla elde edilecek olan veriler bilimsel araştırmalara temel teşkil edeceğinden, bu veri toplama enstrümanlarının hazırlanmasında bazı noktaların dikkate alınması gerekmektedir. Bunlardan bazılarını aşağıda değinilmektedir.

Kategori Sayısı: Ölçeklerde kullanılacak kategori sayısı veya cevap alternatifi sayısına ilişkin birbiri ile çelişen iki durumun dikkate alınması gerekmektedir. Ölçekte kullanılan kategori sayısı arttıkça incelenen konuya ilişkin olarak gruplar arasındaki farklılıkların ayırt edilmesi kolaylaşmaktadır. Başka bir ifade ile, ölçeğin ayırt etme özelliği artmaktadır. Ayrıca kategori sayısı arttıkça elde edilen verilerin normal dağılıma daha fazla yaklaşması (azalan çarpıklık) söz konusu olabilmektedir.³

Öte yandan kategori sayısının artmasıyla birlikte cevaplayıcıların kategoriler arası ayırım yapması zorlaşmaktadır. Dolayısıyla çelişkili bu iki hususun dikkate alınması önemlidir. Geleneksel uygulamalarda kategori sayısının beş ile dokuz arasında olması yönündendir.⁴ Bu konudaki genel tavsiye, cevaplayıcıların ölçekler konusunda tecrübeli olması durumunda çok sayıda kategorinin, tecrübesiz olmaları durumunda ise daha az sayıda kategorili ölçeklerin tercih edilmesi yönündendir.

Bu konudaki diğer bir faktör ise anketin hangi ortamda yapılacağıdır. Telefon aracılığıyla yapılan anketlerde yanlış anlama ve karmaşıklık azaltmak için daha az sayıda kategorinin kullanılması uygun olacaktır. Posta ile yapılan anketlerde ise kategori sayısı nispeten daha az kısıtlayıcı unsur olarak düşünülebilir.⁵

Eğer analiz safhasında, Likert ölçeğinde olduğu gibi, çok sayıda ifadenin toplamı şeklinde ortaya konacak skorlar için beşli ölçek uygun olacaktır. Araştırmanın amacı gruplar arası karşılaştırmalar ve genellemeler yapmak ise beş kategorili ölçek uygun iken, bireysel özellikler ön plana çıkarılmak isteniyorsa kategori sayısının yüksek tutulmasında (yedi veya dokuz) yarar bulunmaktadır. Ayrıca, değişkenler arasındaki korelasyonun da kategori sayısından etkilendiğini dikkate almak gerekmektedir. Kategori sayısı arttıkça korelasyon katsayısı da artma eğilimi göstermektedir.⁶

Ölçeklerde Denge: Ölçeklerdeki olumlu ve olumsuz kategori sayısı da önemlidir. Olumlu ve olumsuz kategorilerin birbirine eşit olduğu ölçeklere *dengeli ölçek*, aksi taktirde *dengesiz ölçek* adı verilmektedir.⁷ Ölçeğin dengeli olup olmamasına bağlı olarak nihai verinin özellikleri de, çarpıklık, etkilenmektedir. Ancak, yapısı gereği deneklerin cevaplarının çarpık olma, (pozitif veya negatif), eğilimi varsa, çarpıklığı azaltmak amacıyla dengesiz bir ölçek tercih edilebilir. Ancak, unutulmaması gereken nokta, veri dağılımının sayıca fazla olan kategori yönünde olacaktır.

Tek veya Çift Sayıda Kategori: Tek sayılı kategorilerin kullanılması durumunda genellikle orta nokta tarafsız veya kararsız seçeneğini temsil eder. Örneğin, Likert ölçeği tek sayıda kategorili bir ölçektir. Tarafsız veya kararsız seçeneğinin varlığı, yeri ve hatta tarafsız etiketinin konması cevaplar üzerinde etkili olmaktadır.⁸ Normal şartlarda tek sayıda kategori içeren ölçeğin kullanılması tercih edilir. Ancak,

arařtırmacı cevaplayıcıları kararsız noktasından uzaklařtırmak istiyorsa veya taraf olmaya zorlamak istiyorsa çift sayıda kategori içeren ölçekler kullanılabilir.

Zorlanmış veya Zorlanmamış Ölçek: Eğer deneğin veya cevaplayıcının konuyla ilgili bir fikri yoksa veya konuyla ilgili olarak fikir beyan etmek istemiyorsa, bu amaca uygun bir *fikrim yok* seçeneđi konabilir. Ancak, arařtırmacı bilerek fikrim yok seçeneđini cevap alternatifleri arasında vermemiş ise, bu ölçeđe *zorlanmış ölçek*, aksi halde *zorlanmamış ölçek* adı verilir. Zorlanmamış ölçekte elde edilen verilerin daha saf olacađı düşünülebilir.⁹

Kategorilerin Yapısı ve Yazılı Gösterimi: Ölçek kategorilerinin doğası ve kelimelerle ifade edilme derecesi cevaplar üzerinde önemli etkiye sahiptir.¹⁰ Ölçek kategorilerini sayısal, kelime veya resimsel olarak göstermek mümkündür. Arařtırmacı ölçekte yer alan bütün kategorileri kelime olarak verip vermeyeceđine karar vermesi gerekmektedir. Özellikle anket konusunda fazla tecrübesi olmayan denekler için bütün kategorilerin isimlendirilmesi yararlı olacaktır. Kategori etiketlerinin ilgili kategoriye en yakın bir yere konması karmařıklığı önleyecektir. Bunun yanında, cevaplayıcıların güçlü argüman içeren (1- kesinlikle katılmıyorum... 7-tamamen katılıyorum gibi) ifadelere daha az cevap verme eğiliminde iken, daha yumuşak olan ifadelere (1-genel olarak katılmıyorum ... 5- genel olarak katılıyor gibi) daha fazla katılma eğiliminde oldukları arařtırmalarda ortaya konmaktadır.¹¹

Ölçeğin Fiziksel Formu: Ölçek formu ve konfigürasyonuna ilişkin çeşitli alternatifler vardır. Kategoriler yatay veya dikey olarak sıralanabileceđi gibi, kategoriler kutucuklar, satırlar, sürekli çizgiler veya pozitif veya negatif değerler şeklinde de ifade edilebilirler. Örneđin, çocuklar üzerinde yapılacak olan bir çalışma için *gülen yüzler* gösterimi oldukça etkilidir.

6.4. Verilerin İnanırcılığı

Veri toplama yöntemi (anket, mülakat veya gözlem) her ne olursa olsun, bilimsel arařtırmaların temel amacı (ve iddiası), arařtırmacının cevabını aradıđı soru veya problemlere doğru (inandırıcı) cevap verebilmektir. Her insan gibi arařtırmacılar da olgulara ve problemlere yaratıcılık yeteneklerine, hayal güçlerine, sezgilerine ve subjektif değerlendirmelerine dayalı olarak izahlar ve çözümler getirirler. Bilimsel yöntem ise, cevabı aranan soruya yanlış cevap alma ihtimalini ortadan kaldırmayı amaçlar. Bunun için yapılan arařtırmalarda iki noktaya özel bir önem verilmektedir: *güvenilirlik* ve *geçerlilik*.

6.4.1. Geçerlilik

Geçerlilik, *bir test veya ölçeğin ölçülmek istenen şeyi ölçme derecesidir*. Bunun için 'ölçekte yer alan sorularla gerçekten ölçmeyi amaçladığımız şeyi ölçebilir miyiz?' sorusunu yöneltmek gerekir.¹² Başka bir bakış açısı ile, geçerliliđi ölçüm değerlerindeki gözlemlenen farklılığın, sistematik veya tesadüfi hatadan kaynaklanmaktan ziyade, ölçülen özellikler açısından nesnel arasındaki gerçek farklılığı yansıtmaya derecesi olarak da tanımlanabilir. Gerçek anlamda geçerliliđe sahip bir ölçüm için sistematik hata ve tesadüfi hata sıfır olmak durumundadır. Ölçümlerde gözlemlenen farklılıklar nesnel arası gerçek farklılıkları yansıttığı düşünülür.

Bu soruya cevap bulmak için deđişik geçerlilik testleri geliştirilmiştir. Bunlardan bir tanesi *içerik geçerliliđi*dir. İçerik geçerliliđi, anketin yeterli sayıda ve ölçülmek istenen olguyu temsil edebilecek soruları içerdikten emin olmak için yapılır. Bir başka deyişle, kavramın öğelerinin ve boyutlarının ölçekte ne derece yer bulduđunun belirlenmesidir. Uygulamada çok deđişik geçerlilik türleriyle karşılaşmaktadır. Yapı geçerliliđi, ölçüt geçerliliđi, öngörüselsel geçerlilik bunlar arasındadır.

Yapısal geçerlilik ölçeğin hangi kavram veya özellikleri ölçtüđünün belirlenmesini incelemektedir. Bu bağlamda arařtırmacı ölçeğin neden doğru olduđunu ortaya koymaya çalışır. Bu amaçla *yakınsak geçerlilik (convergent validity)* ve *ayırtedici geçerlilik (discriminant validity)* kavramlarının uygulanması gerekmektedir. *Yakınsak geçerlilik* sözkonusu ölçeğin aynı yapıyı ölçen diđer ölçütlerle ne derece örtüştüđünün göstergesidir. Aynı yapıyı ölçtüđü düşünölen ölçekler arasında yüksek korelasyonun olması beklenir. *Ayırtedici geçerlilik* ise bir yapıya ilişkin ölçekle farklı yapıları ölçen ölçekler arasında zayıf ilişkinin (düşük korelasyonun) olması anlamına gelmektedir.¹³ Bir ölçeğin geçerliliđinden bahsedebilmek için sözkonusu ölçeğin hem yakınsak hem de ayırtedici geçerliliđe sahip olması gerekmektedir.

6.4.2. Güvenilirlik

Güvenilirlik, 'bir testin veya ölçeğin ölçmek istediği şeyi tutarlı ve istikrarlı bir biçimde ölçme derecesidir'.¹⁴ Güvenilir bir test veya ölçek, benzeri şartlarda tekrar uygulandığında benzeri sonuçlar verir.¹⁵ 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. Zira, aynı test aynı kişilere farklı zamanlarda uygulandığında farklı sonuçlar veriyorsa, elde edilen sonuçlara dayalı yorum yapmak mümkün olamaz. Eğer bir zekâ testinden aynı şahıs bir gün 110, ertesi gün 140 daha sonraki gün 80 alıyorsa bu testin güvenilirliğinden bahsedilemez. Ortada kontrol edilemeyen bazı durumlar var demektir ve bu yüzden elde edilen veriler istikrarlı (güvenilir) değildir.

Bir ölçeğin güvenilirliği söz konusu ölçekte yer alan tesadüfi hatalarla ilgilidir. Ölçeğin yapısında yer alan sistematik hata güvenilirlik üzerinde etkili olmayacaktır. Ölçeğin istikrarını bozan şey random (tesadüfi) hatalardır.

Test ve ölçekler en kolay standart veri toplama ve veri tabanı oluşturma araçlarıdır. Testlerden elde edilen verilerin analiz edilebileceği paket programların (SPSS, SAS, TSP gibi) yaygınlaşması ile birlikte, belli istatistiksel analiz tekniklerinin uygulanması çok kolaylaşmıştır. Buna dayalı olarak Sosyal Bilimlerde giderek daha çok test ve ölçek kullanılmaktadır. Ancak, çok titiz olunmazsa uygulanan bu test ve ölçeklerin güvenilirliklerini azaltacak bir dizi hata yapmak mümkündür.

Bir araştırmamanın güvenilirliğini tespit etmek için iki soruya cevap aramak gerekir:

Kullanılan ölçüm aracı (test, ölçek) farklı şartlar altında benzeri sonuçları veriyor mu?

Farklı araştırmacılar tarafından, farklı durumlarda benzeri gözlemler yapılıyor mu?

Bir ölçeğin güvenilirliğini ölçmede çeşitli yaklaşımlar kullanılmaktadır. Bunlardan birincisi *test-yeniden test* (test re-test reliability) güvenilirliğidir. Buna göre benzer anket formu iki farklı zamanda aynı denekler üzerinde mümkün olduğunca aynı şartlar altında uygulanır. İki ölçüm arasında geçmesi gereken süre ise iki – dört hafta arasındadır. İki farklı zamandaki gözlemler arasındaki korelasyon güvenilirliğin bir göstergesidir. Ancak, bu teste ilişkin çeşitli sorunlar bulunmaktadır. Bunlar arasında bu yaklaşımın zamana karşı olan hassasiyeti, ilk ölçüm deneklerde şartlanmaya sebep olabilme ihtimali, tekrar etme ölçümlerinin zorluğu, halo (önceki sonuçların sonrakisini etkilemesi) etkisi, ölçülen özelliğin ölçekten ölçeğe değişmesi vb. sayılabilir.

İkinci yaklaşım ise *alternatif formlar* yaklaşımıdır. Bu uygulamada aynı anketin eşdeğer iki formatı geliştirilir ve iki-dört hafta arayla aynı deneklere uygulanır. Bu iki ölçüm arasındaki korelasyon güvenilirliğin bir göstergesidir.

Üçüncü yaklaşım ise *içsel tutarlılık* analizidir. Bu yaklaşımda kavramı ölçerken çok sayıda maddeden oluşan, Likert, bir ölçek kullanılarak, ölçekte yer alan maddeler arasındaki korelasyon değerine bakarak içsel tutarlılık analizi yapılır. İçsel tutarlılığı ölçmede iki yaklaşım yaygın olarak kullanılmaktadır. Birincisi, *ikiye ayırma güvenilirliği* (*split-half reliability*) olup, ölçekte yer alan maddeler tesadüfi olarak ikiye ayrılır. Daha sonra bu iki yarının skorları arasındaki korelasyon incelenir. Yaygın uygulamada tek ve çift sayılı maddeler ayrılarak iki grup oluşturulur. Ancak, bölme işlemine bağlı olarak sonuçlar da değişebilmektedir.

Güvenilirliği sorgulamak için anahtar soru:

Eğer araştırma farklı araştırmacılar tarafından, aynı yöntem kullanılarak tekrar yapılırsa aynı/benzeri sonuçlar elde edilebilir mi?

Cevap 'evet' ise araştırma güvenilirdir.

İçsel tutarlılığın ölçümünde en yaygın kullanılan yöntem *Cronbach alfa* olarak da bilinen *alfa katsayısı*dır. Alfa değeri olası tüm ikiye ayırma kombinasyonları sonucu ortaya çıkacak olan ikiye ayırma katsayılarının bir ortalamasını göstermektedir.¹⁶ Alfa değeri 0 ile 1 arası değerler alır ve kabul edilebilir bir değer en az 0.7 olması arzu edilir. Ancak, inceleme türü çalışmalarda bu değer 0.5'e kadar makul kabul edilebileceği de bazı araştırmacılarca öngörülmektedir. Ancak, alfa katsayısının önemli bir özelliği olan ölçekte yer alan madde sayısı arttıkça, alfa değerinin de artmasıdır. Bu durum suni bir artış

olabileceğinden yanıltıcı sonuçlar doğurabilir. Bu konuda yardımcı olacak araç ise beta katsayısının hesaplanmasıdır.

Güvenilirlik ve geçerlilik arasında yakın bir bağ vardır. Eğer bir ölçek mükemmel bir geçerliliğe sahipse, mükemmel bir güvenilirliğe de sahiptir. Eğer ölçek güvenilir değil ise, mükemmel bir geçerliliğe de sahip olamaz, çünkü ölçülen değerler tesadüfi hata terimini de içermektedir. Kısacası, güvenilir olmama geçerliliği de riske sokmaktadır. Ancak, bir ölçeğin güvenilir olması geçerliliğin sağlandığı anlamına gelmeyebilir. Geçerlilik için güvenilirlik bir gerek şarttır, ancak yeterli değildir.

6.5. Ölçmede Hata Kaynakları

Araştırma sürecinde yapılan gözlemler ve ölçümler hata unsurlarını da içermektedirler. Bu ölçüm hatalarının çeşitli kaynakları olabilir. Bazıları anket formu ve soru ölçeğinin tasarımından ve araştırma dizaynında gözden kaçırılan bir nokta sebebiyle tüm ölçümlerde kendini yineleyen hata kaynaklarıdır ki, bunlara sistematik hata denmektedir. Ancak bunun yanında tesadüfi olarak çeşitli öngörülemeyen unsurlardan kaynaklanan tesadüfi (random) hatalar da bulunmaktadır.

Hataların araştırmanın güvenilirliğini tehdit eden değişik faktörler vardır. Bunlardan birincisi *denek hatasıdır*. Sözelimi, çalışanların işlerine bağlılık ve istek derecelerini ölçmeye yönelik bir anket uygulanıyor olsun. Eğer, anket haftanın değişik günlerinde doldurulmuşsa farklı cevapların verileceği açıktır. Pazartesi sabahı ile Cuma öğlen sonrası verilecek cevaplar değişik olur. Böyle bir çalışma için araştırmacı, mümkün olduğunca nötr bir zaman seçmelidir.

Bir başka tehdit ise *denek önyargısıdır*. Özellikle merkezi ve otokratik yapıları organizasyonlardan seçilmiş veya kendini güvende hissetmeyen denekler, sorulara genellikle gerçekte düşündükleri şekilde değil de üstünü memnun edecek şekilde cevap verme eğilimindedirler. Araştırmacı, bu riskin farkında olmalı ve bu tür bir önyargıyı en aza indirmek için deneklerin kimliklerini ortaya çıkaracak sorulardan kaçınmalıdır.

Gözlemci hatası, güvenilirliği tehdit eden bir diğer unsurdur. Örneğin, araştırmacı veri toplama metodu olarak mülakatı seçmiş olsun. Böyle bir çalışmada ne kadar değişik sayıda mülakatçı varsa o kadar değişik cevap almak mümkündür. Bu tehdidi ortadan kaldırmak için, ya mülakatçı sayısını en aza indirmek veya mülakatları mümkün olduğunca biçimselleştirmek gerekir.

Gözlemci önyargısı ise araştırmacının elde ettiği cevapları deneğin kastettiğinden farklı şekilde yorumlamasından kaynaklanır. Gözlemcide deneğin bir şekilde 'takiyye' yaptığı inancı oluşur.

Bölüm Özeti

- Araştırmacılar verilere bir değer atamak için onlara ilişkin ölçümler yaparlar. Ancak, her ölçek (soru biçimi) aynı düzeyde veri edinme imkanı sağlamaz. Bu bakımdan ölçekleri sözde, sıralı, aralı ve oranlı olmak üzere dört gruba ayırmak mümkündür.
- Sözde ölçekleri sadece gruplama yapma imkanı sağlar ancak, bir önemli-önemsiz sıralaması yapmaya uygun değildir. Sıralı ölçeklerde ise, bir astlık-üstlük veya önemli-önemsiz ayrımı, yani sıralama yapmak mümkündür. Ancak, sıralar arasındaki farka dair bilgi vermezler. Aralı verilerde ise hem sıralama yapılmakta hem de fark bilinmemekte ancak, farkın ifade edildiği değerler 'atfedilmekte' gerçek olmamaktadır. Son ölçek türü ise, oranlı ölçeklerdir.
- Bir araştırmanın sonuçlarının inandırıcı olması için öncelikle veri edinme yönteminin inandırıcı olması gerekir. Burada inandırıcılığın iki temel ölçütü olduğu görülmektedir: geçerlilik ve güvenilirlik.
- Geçerlilik, bir ölçüm aracının (sözelimi anketin) veya sorunun ölçmek istenilen şeyi ölçebilme derecesidir. Bu yüzden fizik sınavlarında fizikle ilgili, edebiyat sınavlarında ise şiir ve romanla ilgili sorulara yer verilir. Bir konuda geçersiz olan ölçüm aracı başka bir konuda geçerli olabilir.
- Güvenilirlik ise, ölçüm aracının aynı evrende farklı zaman ve şartlarda uygulanması halinde dahi benzeri sonucu vermesidir. Yani, ölçümün ve elde edilen sonuçların dışsal faktörlerden etkilenmemesidir.

Kaynakça

1. MALHOTRA, N.K. (1996), *Marketing Research – An Applied Orientation*, 2. Edition, Prentice Hall.

2. DICKSON, J. Ve G. ALBAUM (1977), "A method for developing tailor made semantic differentials for specific marketing content areas", *Journal of Marketing Research*, Vol. 21, s.463-71.
3. HAIR, Jr J.F., R.E. ANDERSON, R.L. TATHAM ve W.C. BLACK (1998), *Multivariate Data Analysis*, Fifth Edition, Prentice Hall; Churchill, Jr., G.A. (1987) *Marketing Research – Methodological Foundations*, Fourth edition, Dryden press.
4. COX III, E.P. (1980), "The optimal number or response alternativesfor a scale: A review", *Journal of Marketing Research*, Vol. 17, November, s. 407-22; Reynolds, F.D. ve J. Neter (1982), "How many categories for respondent classification", *Journal of the Market Research Society*, October, s.345-46.
5. MALHOTRA, N.K. (1996), *Marketing Research – An Applied Orientation*, 2. Edition, Prentice Hall.
6. GIVON, M.M. ve Z. SHAPHIRA (1984), "Response to rating scales: A theoretical model and its application to the number of categories problem", *Journal of Marketing Research*, November, s. 410-19.
7. CHURCHILL, Jr. G.A. (1998), *Marketing Research – Methodological Foundations*, Prentice Hall.
8. SPAGNA, G.J. (1984), "Questionnaires: Which approach do you use", *Journal of Advertising Research*, February-March, s.67-70; Holdaway, E.A. (1971), "Different response categories and questionnaire response patterns", *Journal of Experimental Education*, Winter, s.59.
9. SCHNEIDER, K.C. (1985), "Uninformed response rate in survey research", *Journal of Business Research*, April, s.153-62.
10. FRIEDMAN, HH. Ve J.R. LEEFER (1981), "Label versus position in rating scales", *Journal of the Academy of Marketing Science*, Spring, s.88-92.
11. CHURCHILL, Jr. G.A. ve J.P. PETER (1984), Research design effects on the reliability of rating scales", *Journal of Marketing Research*, November, s. 3365-66.
12. PETER, J.P. (1981), "Construct validity: A review of basic issues and marketing practices", *Journal of Marketing Research*, Vol. 18, May, s.133-45.
13. SPIRO, R.L. ve B.A. WEITZ (1990), "Adaptive selling: Conceptualization, measurement and nomological validity", *Journal of Marketing Research*, Vol. 27, February, s.61-69.
14. PETER, J.P. (1979), Reliability: A review of psychometric basics and recent marketing practices" *Journal of Marketing Research*, Vol. 16, February, s.6-17.
15. PERREAULT, Jr. W.D. ve L.E. LEIGH (1989), "Reliability of nominal data based on qualitative judgements", *Journal of Marketing Research*, Vol. 25, May, s.135-48.
16. PETERSON, R.A. (1994), A meta analysisi of Cronbach's coefficinets of alpha"*Journal of Consumer Research*, Vol. 21, September, s.381-91; Cronbach, L.J. (1951), "Coefficient alpha and the internal structure of texts", *Psychometrika*, Vol. 16, s.297-334.