

DAVRANIŐSAL VE SOSYAL BİLİMLERDE YAYGINCA SORGULANAN 10 METODOLOJİK SORUNUN CEVABI *

Responses to the 10 Widely Discussed Methodological Questions among Behavioral and Social Scientists

Volkan Dođan**

Eskişehir Osmangazi Üniversitesi

Özet

Bu metodolojik yazıda, davranıŐsal ve sosyal bilimlerde yaygınca sorgulandıđı düşünölen 10 metodoloji sorusuna cevap sunulmaya çalıŐılmıştır. İlgili metodolojik sorular, kantitatif araştırma metodolojisi alanında psikometrik ölçüm merkezli sorulardır. Dolayısıyla, davranıŐsal ve sosyal bilimlerde psikometrik ölçüm araçlarından faydalanarak kantitatif araştırma tasarımı ile araştırma projeleri yüröten araŐtırmacılar, bu metodolojik yazının hedef okuyucu kitesidir.

Anahtar kelimeler: Metodoloji, Kantitatif, Psikometri, Ölçüm.

Abstract

In this methodological paper, responses to the 10 widely discussed methodological questions among behavioral and social scientists are provided. These methodological questions are measurement-oriented psychometric questions, which have quantitative approach. Thus, researchers who conduct research projects with quantitative methodology in the behavioral and social sciences are the target audience of the current paper.

Keywords: Methodology, Quantitative, Psychometry, Measurement.

* **Makale GeliŐ Tarihi:** 21/08/2017 **Makale Kabul Tarihi:** 17/10/2017

** ArŐ. Gör. Dr. Volkan Dođan, Eskişehir Osmangazi Üniversitesi İİBF, Eskişehir, E-posta: vodogan@ogu.edu.tr

GİRİŞ

...science is not common sense, and its most basic ideas and frames of reference require development through complex intellectual processes which involve not only interpretations of observation but also theoretical and partly philosophical conceptualization.

Talcott Parsons (1968: 429)

Bilimsel arařtırmayı içsel varsayımlardan, önyargılardan ve doğru olarak kanıksanmış fakat test edilmemiş gerçeklik tanımlamalarından ayıran temel özellik sahip olduđu metodolojidir (Neuman, 2002). Metodoloji bir anlamda bilim insanının bilimsel arařtırma yapış şeklini tanımlayan bir terimdir. Bilimsel arařtırmanın yapılaş şekli ise neyin nasıl arařtırıldıđı temeli üzerine kurgulanmaktadır (Bagozzi ve Fornell, 1982). Bilimsel arařtırmanın yapılaş şekli, řu iki soruya arařtırmacının cevabı doğrultusunda şekillenmektedir: (1) Arařtırmacı, arařtırılan fenomenin varlıđını nasıl betimliyor? (2) Arařtırmacı, arařtırılan fenomenin varlıđına iliřkin ne tür bir bilgi ediniminin mümkün olduđunu düşünüyor? Bu sorular sırasıyla ontolojik ve epistemolojik olarak arařtırmaya konu olan fenomene yöneltilen sorulardır (Burrell ve Morgan, 1979). Bu iki soruya verilecek cevaplara bađlı olarak metodolojik tutum iki farklı doğrultuda şekillenebilmektedir. İlki, arařtırılan fenomene iliřkin sadece kendi bađlamı içinde görece cevaplar sunabilmenin mümkün olduđunun düşünöldüđu kalitatif (qualitative) tutumdur (Patton, 2005). İkincisi ise arařtırılan fenomene iliřkin önermelerin sistematik şekilde test edilebilmesinin mümkün olduđunun düşünöldüđu kantitatif (quantitative) tutumdur (Bagozzi ve Fornell, 1982). Bu iki farklı metodolojik tutumdan hangisinin arařtırmacı tarafından benimsendiđi, arařtırma sorularına cevap ararken yararlanacak metodolojik araçların belirleyicisidir.

Hangi metodolojik araçlardan faydalanılması gerektiđi sorusuna cevap bulmanın yanı sıra bu metodolojik araçların kullanılmasının ardındaki teorik gerekçelerin kavranması, metodolojik uygulamalarda yaşanabilecek sorunları azaltacaktır. Metodolojik uygulamalarda

karşılaşılan sorunları giderebilmenin bir yolu akademik dergilerde yayınlanan yönetsel açıdan öğretici yazılardır (örn. methodological teacher corner). Söz konusu kaynakların varlığı her ne kadar Kuzey Amerika merkezli metodoloji dergilerinde gelenek haline gelmiş olsa da Türkiye merkezli akademik dergilerde ilgili uygulamanın yoksunluğu dikkat çekmektedir. İlgili yoksunluk, belirli disiplinlerde çok sayıda araştırmacı arasında tartışılan metodolojik soruların farklı türden cevaplar ile yanıtlanmasına neden olabilme riskini oluşturmaktadır. Bu riski ortadan kaldırmanın yolu yaygınca tartışılan metodolojik sorulara yanıt niteliğinde Türkiye merkezli akademik dergilerde öğretici yazıların yayınlanmasıdır. Bu öğretici yazı ilgili niyetle kaleme alınmıştır.

Bu kapsamda, kantitatif metodolojik tutum ile sınırlı olmak kaydıyla, davranışsal ve sosyal bilimlerden araştırmacılar arasında yaygınca tartışıldığı düşünülen¹ 10 temel metodolojik soruya cevap vermeye çalışılacaktır. Özellikle davranışsal ve sosyal bilimlerde sıkça kullanılan birtakım istatistiksel tekniklere ve yönetsel sorunsallara ilişkin cevapları içeren bu makalenin amacı, Türkiye adresli davranışsal ve sosyal bilimler araştırmacılarını ilgili konular üzerine düşünmeye yöneltmektir.

(1) Kesitsel araştırmalar veya deneysel araştırmalar gerçekleştirilirken örneklem büyüklüğüne nasıl karar verilmelidir?

Araştırma sorusunun belirlenmesi sonrasında araştırma soru(lar)ını test edilebilir kılan araştırma hipotez(ler)i kurgulanır. Kurgulanan hipotez(ler)in test edilebilmesi için ise kesitsel (cross-sectional) araştırmalarda veri toplanmalı ve analiz edilmelidir. Örneğin, materyalizm ile lüks tüketim eğilimi arasındaki ilişkinin test edileceği araştırma projesi için üniversite öğrencilerinden anket aracılığıyla veri toplanması hedeflenebilir. Peki ya örneklem büyüklüğü ne ka-

1 Bu tespit, yazarın kendi akademik deneyimi doğrultusunda ve yakın çevresindeki araştırmacılar ile deneyimlediği düşünsel tartışmalar neticesinde subjektif şekilde yapılmıştır.

dar olmalıdır? Bir diğer deyişle, kaç kişiden veri toplanmalıdır? Bu soru, örneklem sayısının kaç olması gerektiği sorusunu temsil etmektedir. İlgili soruya cevap sunan farklı yaklaşımlar bulunmaktadır. Bu yaklaşımlara değinmek gerekirse: (1) toplanabildiği kadar çok verinin toplanması, (2) belirli bir formül aracılığıyla hesaplanacak olan rakam kadar verinin toplanması (Çil, 2008; Kalaycı, 2010; Kurtuluş, 2010), (3) 384 veya daha fazla sayıda verinin toplanması (Büyüköztürk vd., 2008; Altunışık vd., 2010) önerileri ile ulusal ve uluslararası literatürde karşılaşmak mümkündür. Değinilen üç önerinin de sosyal ve davranışsal araştırmalara yönelik mekanik (pür istatistiksel) yaklaşımın bir sonucu olduğu düşünülmektedir. Özellikle istatistik ve sayısal yöntemler disiplinlerinden araştırmacılardan, davranışsal ve sosyal bilimlerdeki araştırmacıların istatistiksel analizlerinde yardım almaları doğrultusunda bu örneklem hesaplama yaklaşımlarının yaygınlaştığı çıkarımında bulunmak mümkündür.

Değinilen üç örneklem hesaplama önerilerinden ilki olan ‘toplanabildiği kadar çok verinin toplanması’ önerisi araştırma ekonomisi ve araştırma sonuçlarının genelleneceği popülasyon (evren) homojenliği açısından verimsiz bir yaklaşım olduğu düşünülmektedir. Detaylandırmak gerekirse; eğer oldukça homojen bir popülasyona (evrene) sonuçları genelledebilmek için veri toplanıyorsa, çok az sayıda örneklem büyüklüğü sayesinde popülasyona sonuçlar genellenebilir (Gonzalez, 2009). Örneğin, her bir üyesi aynı niteliğe sahip olan bir hayali popülasyondan ister 1 ister 1000 veri toplansın sonuçlar popülasyonu temsil eder nitelikte olacaktır. Bu durumda ‘toplanabildiği kadar çok verinin toplanması’ önerisi ekonomik ve zamansal kaynakların boşa harcanmasına neden olacaktır. Dolayısıyla araştırmacı örneklem büyüklüğü hesaplaması yapmadan önce araştırma sonuçlarını hangi popülasyona genellemeyi amaçladığı üzerine ve bu hedef popülasyonun ne derece homojen olduğu üzerine düşünmelidir. İlgili hedef popülasyonun homojenliği arttıkça, daha küçük örneklem büyüklüğü ile genellenebilir bulgulara ulaşma olasılığı da artma yönünde eğilim sergilemektedir. Bir diğer deyişle, oldukça heterojen bir hedef popülasyon için örneklem büyüklüğü, oldukça homojen bir

hedef popülasyon için örneklem büyüklüğünden daha fazla olmalıdır (Rosnow ve Rosenthal, 1996). Bu noktada arařtırmacı davranıřsal ve/veya sosyal bir arařtırma yürüttüğü bilincinde olup popülasyonun karakteristiđini önceden göz önünde bulundurmalıdır.

Deđinilen üç örneklem hesaplama önerilerinden ikincisi olan ‘belirli bir formül aracılıđıyla hesaplanacak olan rakam kadar verinin toplanması’ önerisinin davranıřsal veya sosyal arařtırmanın dođasına aykırı mekanik (pür istatistik) yaklařımın bir ürünü olduđu düşünölmektedir. Sihirli bir rakam hesabı yaparcasına uygulanan örneklem büyüklüğü formöllerini, ulusal alanda yaygın etkiye sahip Türkçe yöntem kaynaklarında da yer bulmuřtur (Çil, 2008; Kalaycı, 2010; Kurtuluş, 2010). Bu durumun başlıca sebebi, davranıřsal ve sosyal bilimler özelindeki arařtırma yöntemleri ile fen bilimleri özelindeki arařtırma yöntemlerini aynı şekilde ele alma eğilimidir. Bu eğilimin, istatistik ve sayısal yöntemler alanından arařtırmacıların pür pozitivist yaklařımla davranıřsal ve sosyal bilimler arařtırmacılarının verilerini analiz etme pratiđinden beslendiđi düşünölmektedir. Örneklem büyüklüğü hesaplamaya yönelik kullanılan formöller; popülasyonun büyüklüğünün net şekilde belli olduđu, popülasyon varyansının net şekilde belli olduđu veya örneklem varyansının varsayımsal olarak belirlendiđi formöllerdir. Dahası, bu formöller popülasyonun karakteristiđini normal dađılım sergilediđi yönünde varsayımlamaktadır. Fakat her bir bireyine ulařılamayan popülasyonun normal dađılım sergilediđini varsaymak ne derece sađlıklıdır? Üstelik hangi deđişken dođrultusunda popülasyonun normal dađılım sergilediđi varsayılmaktadır? Bu iki soruya, formül dođrultusunda örneklem büyüklüğü hesaplama yaklařımı sađlıklı cevap üretmekte yetersiz kalmaktadır. Dolayısıyla ‘belirli bir formül aracılıđıyla hesaplanacak olan rakam kadar verinin toplanması’ yaklařımı ile örnekleme büyüklüğü belirleme önerilmemektedir. Davranıřsal ve sosyal bilimler kapsamındaki arařtırma projeleri büyük ölçüde insanlardan oluřan bir popülasyonu ilgilendirmektedir. Arařtırmanın yapıldıđı an itibariyle popülasyonun net şekilde belirlenmesi neredeyse imkansızdır. Bu yüzden bazı metodologlar davranıřsal ve sosyal bilimlerde popülasyon yeri-

ne örneklem çerçevesi (sampling frame) kavramını kullanmaktadırlar (Neuman, 2002).

Değınilen üç örneklem hesaplama önerilerinden üçüncü ve sonuncusu olan '384 veya daha fazla sayıda verinin toplanması' önerisi, bir önceki öneri olan formül aracılığıyla örneklem büyüklüğü hesaplama önerisinin bir uzantısı niteliğindedir. Belirli güven aralıkları ve anlamlılık seviyeleri önceden belirlenerek hesaplanan 384 sihirli rakamının, popülasyon büyüklüğü ne derece büyük olursa olsun yeterli örneklem sayısı olarak savunulduğu ulusal ve uluslararası kaynakların bulunması (Barlett vd., 2001; Büyüköztürk vd., 2008; Altunışık vd., 2010) sıkıntılı bir durumdur. Davranışsal ve sosyal bilimlerde bilimsel araştırma yapma sürecine bu derece mekanik yaklaşım, araştırmacıyı tasarım üzerine düşünmeden uzaklaştırma riskini beraberinde getirebilir. Bu yüzden mekanik doğaya sahip örnekleme büyüklüğü hesaplamaları pratikleri terk edilmeli ve popülasyonun karakteristiğine uygun örnekleme büyüklüğü belirleme yaklaşımı benimsenmelidir.

Daha önceki paragraflarda değınilen örneklem büyüklüğü belirleme önerilerinin aslında araştırmacının sihirli bir örneklem büyüklüğü rakamı elde edebilme arzusu doğrultusunda meydana geldiği savunulabilir. Oysa örneklem büyüklüğü meselesi metodolojik anlamda araştırmanın doğası yani araştırmacının temel amacı ile ilgilidir. Bu doğrultuda örneklem büyüklüğü hesaplaması yapılması bir zorunluluk ise testin gücü (power of test) yaklaşımının tercih edilmesi önerilmektedir. Nitekim testin gücü, alternatif hipotezin doğru olduğu durumda boş hipotezin reddedilebilme olasılığıdır (Hair vd., 1998). Metodoloji literatüründe önemli tartışmalara neden olan (Cohen, 1992; MacCallum vd., 1996; Faul vd., 2009; Simonsohn, 2015) testin gücü meselesi, özellikle hipotez testine odaklanan bir yaklaşım sunmaktadır. Bu yüzden Tip-1 (yanlış şekilde boş hipotezin reddi) ve Tip-2 (yanlış şekilde boş hipotezin reddedilememesi) hataların belirli değerlerde tutulması, testin gücü yaklaşımının merkezinde yer almaktadır (Aronson vd., 1990; Rosenthal ve Rosnow, 1991). Testin

gücü yaklaşımının odak noktasında hipotez testi yer alırken daha önce değinilen diğer üç farklı örneklem büyüklüğü hesaplama yaklaşımının odak noktasında ise popülasyona ilişkin olası parametreler yer almaktadır. Dolayısıyla her ne kadar tüm bu yaklaşımlar istatistiksel hesaplamaları bünyesinde barındırsa da testin gücü yaklaşımında popülasyon ilişkin olası parametre değerleri varsayımı söz konusu değildir.

Bu yazı kapsamında önerilen testin gücü yaklaşımı doğrultusunda örneklem büyüklüğü hesaplamayı benimsemeden önce araştırmacı kendisine şu soruyu sormalıdır: Bu araştırmayı yapmanın temel sebebi bir teori testi gerçekleştirmek mi? yoksa ulaşılan bulguları belirli bir popülasyona genelleyebilmek mi? Örneğin; Türkiye’de son 10 yılda pazarlama doktorasını tamamlamış kimselerin doktora sonrasında yaptıkları yayın sayısının doktora sürecinde araştırma yöntemi dersi alıp almamış olmaları ile ilişkili olup olmadığı sorusuna cevap arayan bir araştırmacı, teori testi gerçekleştirme değil belirli bir popülasyona genelleyebilmek üzere bazı bulgulara ulaşmak amacıyla bu araştırmayı gerçekleştirmektedir. Öte yandan, tüketicinin kurgusal seviyesi (construal level theory, Trope ve Liberman, 2010) ile ürünün hedonik mi yoksa fonksiyonel bir ürün mü olduğu bileşiminin satın alma niyetini belirlediği hipotezinin test edildiği bir deneysel araştırmada ise amaç, sonuçları popülasyona genellemek değil teori testi gerçekleştirmektir. Dolayısıyla, davranışsal ve sosyal bilimler kapsamında araştırmalar gerçekleştiren bir araştırmacı, araştırma projesinde teori testi gerçekleştirmeye odaklanıyor ise hipotez(ler)ini test etmek için yeterli test gücünü sunan örneklem büyüklüğünün üzerine çıkmak kaynak israfıdır. Bu doğrultuda testin gücü hesabı yapılarak yeterli gücü (önerilen testin gücü %80; Cohen, 1992) sunan örneklem büyüklüğü belirlendikten sonra veri toplama süreci gerçekleştirilmelidir.

Peki ya araştırmacının ana amacı araştırma projesi kapsamında ulaşacağı bulguları hedef popülasyona genelleyebilmek ise örneklem büyüklüğü nasıl belirlenmeli? Bu sorunun cevabı, araştırmacı-

nın olasılığa dayalı örnekleme yöntemini mi yoksa olasılığa dayalı olmayan örnekleme yöntemini mi tercih edeceğine bağlıdır. Eğer olasılığa dayalı örnekleme yöntemi tercih edilecek ise hem testin gücü hesaplaması neticesinde ulaşılabilecek rakamdan hem de popülasyon büyüklüğünün hesaba katıldığı örnekleme büyüklüğü hesaplama formülü neticesinde elde edilen rakamdan büyük olacak şekilde örneklem büyüklüğü sayısı belirlenmelidir. Eğer araştırmacı olasılığa dayalı olmayan örnekleme yöntemi doğrultusunda veri toplayacak ise, herhangi bir örneklem büyüklüğü hesaplama önerisinde bulunulmamaktadır. Çünkü yapılacak olan herhangi bir hesaplamanın örneklemin popülasyonu temsiline ilişkin söyleyebileceği hiçbir şey yoktur. Bir diğer deyişle, örneklemden elde edilen bulguların popülasyona genellenebilirliğine ilişkin bir hesaplama söz konusu değildir. Ancak popülasyon ne kadar homojen bir karakteristiğe sahip ise araştırmacının örneklemden elde edeceği bulguyu popülasyona genelleyebilme olasılığı o kadar yüksektir. Ancak bu olasılığa ilişkin bir hesaplama yapabilmek mümkün değildir. Bu noktada araştırmacının ekonomik kaynağı oldukça iyiye 'toplanabildiği kadar çok verinin toplanması' önerisinin takip edilmesi düşünülebilir.

Öte yandan 'toplanabildiği kadar çok verinin toplanması' önerisini takip etmenin neden olacağı en önemli sorun p-değerinin örneklem büyüklüğüne karşı hassas olmasıdır. Bir diğer deyişle, örneklem büyüklüğü arttıkça p-değeri (istatistiksel anlamlılık testi) küçülme eğilimi göstermektedir (Simonsohn vd., 2014). Dolayısıyla araştırmacı 200 veri topladıktan sonra istatistiksel olarak anlamsız bulunduğu ilişkiyi ($p = .15$) 300 veri daha topladıktan sonra istatistiksel olarak anlamlıya dönüştürebilir ($p = .01$). Bu yüzden 'toplanabildiği kadar çok verinin toplanması' önerisi Tip-1 olasılığını yanlış şekilde arttırmaktadır. Sonuç itibarıyla, genellenebilir fakat istatistiksel olarak yanlış bulgulara ulaşmak 'toplanabildiği kadar çok verinin toplanması' önerisi ile kaçınılmaz olmaktadır.

Bu metodolojik yazı kapsamında önerilen örneklem büyüklüğü hesaplama yaklaşımı olan testin gücü yaklaşımına detaylı değinmek

gerekirse, testin gücü hesabının veri toplanmadan önce veya veri toplandıktan sonra olmak hesaplanabildiğini bilmek gerekir. Örneğin, materyalizm ile lüks tüketim eğilimi arasındaki ilişkiyi test etmek üzere psikometrik ölçekler aracılığıyla 250 veri toplandıktan sonra testin gücü hesaplanabilmektedir. Bu hesaplamayı G*Power (Faul vd., 2009) programı kullanarak araştırmacılar yapabilmektedirler. İlgili program sayesinde istenildiği zaman testin gücü hesabı yapılabiliyor olmasına rağmen, testin gücü hesabının veri toplanmadan önce yapılması önerilmektedir (Cohen, 1992). Çünkü bilindiği üzere testin gücü, yanlış olan boş hipotezin red edilebilme olasılığıdır (Cohen vd., 2003). Bir diğer deyişle, materyalizm ile lüks tüketim eğilimi arasında gerçekten bir ilişki olduğu durumda biz boş hipotez olan materyalizm ile lüks tüketim eğilimi arasında ilişki yoktur hipotezini ne olasılıkla red edebiliyorsak o düzeyde testin gücüne sahibiz demektir. Dolayısıyla veri toplandıktan sonra yapılacak testin gücü hesaplamasının düşük olması durumunda, araştırmacı sadece testin zayıflığının farkına varabilir fakat bu durumu düzeltmek için fazladan veri toplanması arzu edilen bir yaklaşım değildir. Oysa araştırmacı veri toplama sürecine başlamadan önce testin gücünü önerilen %80 veya daha yüksek (Cohen, 1992) olacak şekilde belirleyip örneklem büyüklüğü hesabı yaptıktan sonra veri toplamış olsa daha yerinde olacaktır. Üstelik veri toplamadan önce testin gücü yapılarak bu doğrultuda örneklem büyüklüğü hesaplamak araştırmacının gereğinden fazla veri toplaması sorununu da ortadan kaldıracaktır. Ayrıca G*Power programını araştırmacılar <http://www.gpower.hhu.de/en.html> adresinden ücretsiz edinebilirler.

(2) Genellenebilir niteliğe sahip bir örneklem her zaman öğrenci örnekleminden daha mı iyidir?

Davranışsal ve sosyal bilim araştırmacıları yaşadıkları zamansal ve ekonomik kısıtlardan dolayı üniversite öğrencisi örneklemelerini sıklıkla kullanmaktadırlar (Espinosa ve Ortinau, 2016). Hatta yönetim bilimleri ve pazarlama disiplinlerindeki görgül bilgi birikiminin çoğunun üniversite öğrencilerinin katılımcılar olduğu araştırmalardan

oluştığı ve bu durumun dışsal geçerliliği (external validity) tehdit ettiği için sürdürülmemesi gerektiği savunmaktadır (Enis vd., 1972; Ferber, 1977; Gordon vd., 1986). Bu yaklaşımın karşısında yer alan araştırmacılar ise üniversite öğrencilerinin katılımcı olarak kullanılmasının sorunlu bir durum olmadığını hatta arzulanan bir durum olduğunu savunmaktadır (Berkowitz ve Donnerstein, 1982; Calder vd., 1982; Lynch, 1982; Lynch, 1983; Dobbins vd., 1988; Winer, 1999). Üniversite öğrencisinin katılımcı olarak kullanılmasını ve kullanılmamasını savunan iki kutup arasında yaşanan metodolojik tartışma geçerlilik kavramına ve araştırmanın amacı kavramına odaklanmayı gerektirmektedir.

Üniversite öğrencisi örnekleme ile genellenebilir niteliğe sahip örneklem arasında bir tercihte bulunmak zorunda kalındığında araştırmanın amacına göre bir tercih yapılmalıdır. Üstelik bazı durumlarda üniversite öğrencisi örnekleminin genellenebilir niteliğe sahip örnekleme karşı üstünlüğü olduğu bilinmektedir. Bu tartışmayı daha derine götürebilmek için ilk olarak etki araştırması (effects application) ile teori araştırması (theory application) ayrımı (Calder vd., 1981) netleştirilmelidir. Bu ayrıma göre; etki araştırması, araştırmacının gözlemlediği bulguları genellemeyi amaçladığı araştırma türü iken teori araştırması ise araştırmacının teorik testleri gerçekleştirmeyi amaçladığı araştırma türü olarak kavramsallaştırılmaktadır (Calder ve Tybout, 1999). Etki araştırmalarında katılımcıların gerçek yaşam koşullarına oldukça yakın çevrede ve popülasyonu temsil edecek nitelikte olması arzulanan iken teori araştırmasında katılımcıların oldukça homojen olması arzulanmaktadır. Dolayısıyla etki araştırması gerçekleştiren bir araştırmacı için genellenebilir niteliğe sahip bir örneklem ile çalışmak ideal iken teori araştırması gerçekleştiren bir araştırmacı için homojen bir örneklem ile çalışmak idealdir. Peki ya bu farklılık neden kaynaklanıyor?

Etki araştırması için en uygun metodolojik yöntem anket araştırmaları iken teori araştırmaları için ise en uygun metodolojik yöntem deneysel araştırmalardır. Çünkü kesitsel araştırmaların temelini

örneklemden elde edilen bulguların popülasyona genellenebilirliği oluşturmaktadır. Deneysel araştırmaların temelini ise içerik geçerliliği yüksek şekilde bağımsız değişken(ler) ile bağımlı değişken(ler) arasındaki nedensellik ilişkilerinin ortaya çıkarılması oluşturmaktadır (Gonzalez, 2009). Ayrıca anket araştırmalarında dışsal geçerlilik ön planda tutulurken deneysel araştırmalarda ise içsel geçerlilik ön planda tutulmaktadır. İçsel geçerlilik, bağımlı değişkendeki değişimin sadece ve sadece bağımsız değişkenden kaynaklandığını ortaya koyabilme olarak ifade edilebilir. Dışsal geçerlilik ise bulguların farklı bağlamlara ve popülasyonlara genellenebilirliği olarak ifade edilebilir (Aronson vd., 1990).

Önceki paragraflardan çıkarılabileceği üzere etki araştırmalarında genellenebilir niteliğe sahip örneklemin kullanılması tercih edilmelidir. Çünkü temel amaç bulguları hedef popülasyona genelle-mektir. Ancak genellenebilir niteliğe sahip olan heterojen bir örneklem ile çalışırken elde edilecek bulguların içsel geçerliliği büyük tehlike altındadır. Bu vurguyu örneklendirmek gerekirse; materyalizm ile marka bağlılığı arasındaki ilişkinin incelendiği ve 5000 tüketiciden oluşan oldukça heterojen bir örneklem kapsamında analizlerin gerçekleştirildiği araştırma neticesinde ulaşılan bulgunun ardında gizli ılımlaştırıcı (moderator) değişkenlerin rolü söz konusu olabilir. Detaylandırmak gerekirse; örnekleme yer alan 5000 kişinin her gelir grubundan tüketiciyi içerdiği ve ulaşılan bulgunun materyalizm ile marka bağlılığı arasında anlamsız ilişki ($r = 0.03$, $p = 0.96$) tespit edildiği varsayılırsa, bu ilişkinin düşük gelirli tüketiciler kapsamında pozitif anlamlı ilişkili ($r = 0.56$, $p = .02$) iken yüksek gelirli tüketiciler kapsamında negatif anlamlı ilişkili ($r = -.45$, $p = .03$) olduğu için genel örneklem kapsamında anlamsız ilişki bulunmuş olabilir. Bu durumda gizli ılımlaştırıcı (moderator) değişken gelir düzeyidir. Bu noktada gelir düzeyinin ılımlaştırıcı (moderator) değişken olduğunun tespit edilmesi sınırlılık şartının (boundary condition) açığa çıkarılması olarak da kavramsallaştırılmaktadır (Wansink ve Ittersum, 2016). Toparlamak gerekirse, ulaşılan anlamsız ilişki bulgusu materyalizm ile marka bağlılığı arasındaki ilişkide gelir düzeyinin

etkisinin göz ardı edilmesi sonucunu önermektedir.¹ İşte bu noktada marka bağlılığındaki değişim sadece materyalizm kaynaklı değil aynı zamanda gelir düzeyi kaynaklı gerçekleşmektedir. Bu yüzden içsel geçerlilik oldukça düşüktür. Ancak içsel geçerlilikten fedakarlık etmeden dışsal geçerliliği artırabilmek pek mümkün olmadığından, etki araştırmalarında dışsal geçerlilik uğruna içsel geçerlilikten bir anlamda fedakarlık edilmektedir.

Teori araştırmalarının için en uygun araştırma yöntemi olan deneysel araştırmalar kapsamında oldukça steril bir şekilde nedensellik ilişkileri ortaya konmaya çalışılmaktadır. Buradaki steril vurgusu içsel geçerliliğin oldukça yüksek tutulması yani bağımlı değişkendeki değişimin sadece ve sadece bağımsız değişkendeki değişimden kaynaklanmasını ifade etmektedir. Deney koşullarına (experimental conditions) rassal şekilde katılımcılar atanarak bu sterillik sağlanmaktadır (Campbell ve Stanley, 1966). Örneğin; parayı düşünmenin materyalizm üzerindeki etkisinin araştırıldığı iki koşuldan oluşan bir gruplararası deneysel tasarımda (between-subject design) deney koşuluna atanan katılımcılar paraya ilişkin bir metin okurken, kontrol koşuluna atanan katılımcılar iklim değişimine ilişkin bir metin okumuş olsunlar. Böylece deney koşuluna atanan katılımcılar parayı düşünmeye sevk edilirken kontrol koşuluna atanan katılımcılar parayla ilgili bir düşünceye sevk edilmemiş olacaktır. Bu manipülasyon sonrasında her bir gruba atanan katılımcılar materyalizm ölçeğini cevaplamış olsunlar. Eğer deney koşulundaki katılımcıların materyalizm ortalaması istatistiksel olarak anlamlı şekilde kontrol koşulundaki katılımcıların materyalizm ortalamasından yüksek tespit edilebilirse parayı düşünmenin materyalizm üzerindeki nedensellik etkisi ortaya konmuş olacaktır. Görüldüğü üzere deney grubu ortalamasından kontrol grubu ortalamasının çıkarılması neticesinde ulaşılan nedensellik sonucu deney grubu ile kontrol grubuna atanmış katılımcıların birbirlerinden ne derece farklı özelliklere sahip (heterojen) iseler o

1 İlgili yanlış sonuç, modelin yanlış kurgulanması veya yetersiz olması durumundan da kaynaklı gerçekleşebilmektedir.

düzeyde tehdit altında olacaktır. Bu yüzden özellikle gruplararası deneysel tasarım uygulamalarında araştırmanın katılımcıların olabildiğince homojen olması araştırmanın bulgularının içsel geçerliliği için kritik öneme sahiptir.

Bir diğer perspektiften deneysel araştırma tasarımı kapsamında heterojen örneklemin sakıncasını açıklamak için sinyal-gürültü (signal-noise) metaforu kullanılabilir. Bilindiği üzere teori testlerinin ön planda olduğu deneysel araştırmaların bulgularının çoğu F-testi neticesinde açığa çıkarılmaktadır. Bu yüzden varyans analizi (ANOVA), deneysel araştırma tasarımının sıklıkla kullanıldığı disiplinlerdeki doktora eğitimlerinde büyük öneme sahiptir. Gruplararası ortalama farklılıklarının testine odaklanan deneysel araştırmalarda uygulanan ANOVA testlerinde gruplararası varyansın (between sum of squares) grup içi varyansa (within sum of squares) oranı neticesinde F değeri hesaplanmaktadır (Gonzalez, 2009). Bu noktada istatistiksel olarak anlamlı bir sonuç bulabilmek için gruplararası varyansın (between sum of squares) grup içi varyanstan (within sum of squares) daha yüksek olması gerekmektedir. Bazı metodologlar gruplararası varyansı (between sum of squares) sinyal olarak, grup içi varyansı (within sum of squares) ise gürültü olarak ifade etmektedirler (Aronson vd., 1990; Rosnow ve Rosenthal, 1996). Böylece ANOVA testinin sonucu olan F oranı signal-to-noise oranı olarak da ifade edilebilir. Bu kavramsallaştırmada deney kapsamında koşullara atanan katılımcılar arası manipülasyon dışında her türlü farklılık gürültüyü artıran bir unsurdur. Dolayısıyla deneysel araştırmanın katılımcıları ne düzeyde heterojen bir örnekleme oluşturuyorlar ise o düzeyde istatistiksel olarak anlamsız sonuçlar bulma olasılığı yüksek olacaktır. Nitekim anlamlı bir nedensellik var iken sonucu anlamsız olarak tespit etmek Tip-2 hatayı (β) ifade ettiği için ve testin gücü de $1-\beta$ olduğu için deneysel araştırmada heterojen bir örneklem ile çalışmak testin gücünü düşürmektedir. Araştırmacılar için arzulanan bu durumdan kaçınmanın yolu olabildiğince homojen bir örneklem ile çalışmaktır.

Toparlamak gerekirse, genellenebilir niteliğe sahip bir örneklemin mi yoksa üniversite öğrencisi örnekleminin mi daha uygun olduğu

araştırmanın etki araştırması mı yoksa teori araştırması mı olduğuna bağlıdır. Araştırma bulgularının popülasyona genellenmesinin amaçlandığı etki araştırması yapılıyorsa oldukça heterojen yapıya sahip genellenebilir nitelikte örneklem ideal iken teori test(ler)inin yapılmasının amaçlandığı teori araştırması yapılıyorsa oldukça homojen yapıya sahip bir örneklem daha idealdir. Bu yüzden genellenebilir niteliğe sahip bir örneklem, eğer üniversite öğrencisi örnekleme daha homojen ise, her zaman üniversite öğrencisi örneklemeden daha iyi değildir. Hatta teori testi söz konusu olduğunda üniversite öğrencisi örnekleme ile çalışmak birçok araştırmacı tarafından daha arzulanır bir durum olarak değerlendirilmektedir (Lynch, 1982).

(3) İki veya daha fazla sayıdaki farklı kültürden toplanmış olan kesitsel veri nasıl analiz edilmelidir?

Davranışsal ve sosyal bilimlerde araştırmalara konu olan fenomenler genellikle elle tutulamaz ve gözle görülemez niteliktedir. Bir diğer deyişle, davranışsal ve sosyal bilimlerde araştırmalara konu olan fenomenler gizil yapılar (latent constructs) olarak kavramsallaştırılıp ölçümleri ise psikometrik ölçüm araçları olan ölçekler aracılığıyla gerçekleştirilmektedir (MacKenzie vd., 2011). Benzer şekilde, yönetim bilimleri ve pazarlama disiplinlerinde kantitatif (nicel) yaklaşımla araştırmalar gerçekleştiren araştırmacılar da gizil yapılar kurgulayıp bu gizil yapıları ölçekler aracılığıyla psikometrik şekilde ölçümlemeye çalışmaktadırlar. Bu ölçümleme bazen birden fazla kültürden toplanan veriyi kapsayabilmektedir. Kültürlerarası verinin analizi neticesinde ise araştırmacılar kültürlerarası karşılaştırmalar yapabilmeyi amaçlamaktadırlar. Peki ya kültürlerarası veri analizi gerçekleştirirken dikkat edilmesi gereken noktalar nelerdir?

Dikkat edilmesi gereken noktalara değinmeye başlamadan önce yapılan önerilerin ve açıklamaların özellikle anket içerisinde yer alan ölçeklerle yapılan ölçümleri kapsadığı unutulmamalıdır. Davranışsal ve sosyal bilimlerde yürütülen gözleme dayalı (empirical) araştırmaların birçoğunun araştırma modeline sahip olduğu bilinmektedir. Burada adı geçen gözleme dayalı vurgusu aslında direkt olmayan (in-

direct) bir gözlemi ifade etmektedir. Bir diğer deyişle, araştırma modeli içinde yer alan her bir değişkenin ölçüm araçları aracılığıyla gözlemlenmesi davranışsal ve sosyal bilimlerde yaygın araştırma yapış şeklidir. Araştırma modeli içinde yer alan her bir değişken bir fenomen yani bir gizil yapıyı temsil etmektedir. Bu gizil yapıları ölçümlemek için kullanılan ölçüm araçları olan psikometrik ölçekler ise ilgili yapılara ilişkin sayısallaştırılabilir ölçümler yapabilmeyi mümkün kılmaktadır (DeVellis, 2017). Soyut yapılar olan gizil yapıların operasyonelleştirilebilmesi için izlenen yol genellikle ölçekler aracılığıyla cevaplayıcıların rapor ettiği cevaplardır. Örneğin; materyalizm değişkeni veya gizil yapısını ölçümleyebilmek adına Richins ve Dawson (1992) tarafından geliştirilmiş olan materyalizm ölçeği ifadelerine cevaplayıcıların sunduğu cevaplar kullanılabilir. Bu noktada, materyalizm ölçeği, soyut yapı olan materyalizm değişkeninin somut uzantısı olarak kavramsallaştırılabilir. Soyut doğaya sahip olan gizil yapıların somut uzantıları tarafından ne düzeyde kapsayıcı şekilde ölçümlenebildiği oldukça öneme sahiptir. Nitekim bu soyut yapı ile somut ölçümün örtüşmesi ‘uyum kuralı’ (correspondence rule) olarak da isimlendirilmektedir (Bagozzi, 2011). Peki ya ölçek ifadeleri aracılığıyla yapılan somut ölçüm ölçümlenmeye çalışılan soyut yapıyı ne düzeyde temsil ediyor sorusunun cevabına nasıl ulaşılabilir?

Bu sorunun cevabını, klasik test teorisinin temel varsayımı sunmaktadır (Nunnally, 1978): Gözlemlenen skor (observed score) = Doğru skor (true score) + Hata (error). Bu denklem, ölçekler aracılığıyla yapılan gözlem neticesinde elde edilen verinin içinde hem ölçülmesi planlanan yapıya ilişkin ölçümün hem de hatanın var olacağını ileri sürmektedir. İlgili denklem şu örnek aracılığıyla da zihinlerde kavramsallaştırılabilir; düşünün ki pazardan 100 gram pirinç satın alarak evinize döndünüz. Eve döndükten sonra pirinç içerisinde yer alan taşlar hataları temsil ederken geriye kalan sağlam pirinçler ise doğru skoru temsil edecektir. Klasik test teorisinin temel varsayımı olan denklem biraz daha detaylandırılırsa, ölçekler aracılığıyla yapılan ölçümlerin üç bileşenden oluştuğu ileri sürülmektedir. Bir diğer deyişle: Ölçüm = Yapı varyansı (trait variance) + Yöntem kaynaklı varyans

(method variance) + Hata kaynaklı varyans (error variance) denklemini aracılığıyla psikometrik ölçümler kavramsallaştırılabilir (Jarvis vd., 2003). Bu denklemdeki unsurları açıklamak gerekirse; yapı varyansı (trait variance) ölçümlenmek istenen gizil yapıya ilişkin yapılan ölçümü temsil ederken, yöntem kaynaklı varyans (method variance) veri toplama sürecinde kullanılan yöntemden kaynaklanan varyansı temsil etmektedir. Hata kaynaklı varyans (error variance) ise yapılan ölçüm içerisinde yer alan hatayı temsil etmektedir.

Davranışsal ve sosyal bilimlerdeki araştırmacıların ölçüm araçları aracılığıyla ölçüm yapmalarının sebebi, araştırma model(ler)inde veya hipotez(ler)inde yer alan gizil yapıları (değişkenleri) ölçümlenebilmektir. Dolayısıyla yapılan ölçüm içerisinde yer alan yapı varyansı (trait variance) araştırmacılar için önemli olan bileşendir. Nitekim iki farklı kültürden ölçek(ler) aracılığıyla toplanmış olan verinin karşılaştırılması yapılırken de iki farklı veri içindeki yapı varyansları (trait variance) karşılaştırılmalıdır. Daha net bir ifadeyle, kültürlerarası karşılaştırmaların amacı sadece ölçek ortalamalarının karşılaştırılması değil ölçeğin ölçmeye çalıştığı gizil yapının (değişkenin) karşılaştırılmasının yapılmasıdır (Steenkamp ve Baumgartner, 1998). Bu nedenle ölçüm araçları aracılığıyla toplanan veri(ler)nin içinden yöntem kaynaklı varyans (method variance) ile hata kaynaklı varyans (error variance) ayıklandıktan sonra kültürlerarası karşılaştırmalar yapılmalıdır.

İki veya daha fazla sayıda grup ortalaması karşılaştırması (*t* testi, *F* testi) ulusal ve uluslararası araştırmacılar tarafından sıklıkla uygulanan analizlerdir. Bu karşılaştırmalarda grupları oluşturan unsur bazen cinsiyet, bazen aylık gelir olabilir iken bazen de farklı kültür veya ülkeden toplanmış veri de olabilmektedir. Örneğin; materyalizm ölçeği (Richins ve Dawson, 1992) aracılığıyla yapılan materyalizm ölçümü neticesinde elde edilen Türkiye verisi ortalaması ile Malezya verisi ortalaması direkt şekilde *t*-testi uygulanarak karşılaştırıldığında elde edilen sonuç materyalizm gizil yapısının kültürlerarası karşılaştırmasına ilişkin sakıncalı olacaktır. Çünkü elde edilen istatistiksel

anlamli veya anlamsız sonucun, verinin bileşenleri olan yapı varyansına (trait variance) mı, yöntem kaynaklı varyansa (method variance) mı, yoksa hata kaynaklı varyansa (error variance) mı atfedileceği bilinemez bir cevaptır. Bir diğ er deyiş le, iki kültürün materyalizm ortalamaları arasındaki fark, yöntem kaynaklı varyansın gruplararası oldukça farklı olmasından kaynaklı olabilir ve araştırmacı bu farkı materyalizm yapısının iki kültürde farklı olmasına dayandırabilir. Böyle bir çıkarım yanlı (biased) ve hatalı bir çıkarım olacaktır. Bu sorunun önüne geçebilmek için iki veya fazla sayıdaki kültürden toplanan verinin karşılaştırması yapılırken ilk adım olarak ölçüm eş varyanslılığı (measurement invariance) test edilmelidir (Steenkamp ve Baumgartner, 1998; Bagozzi, 2011; MacKenzie vd., 2011). Peki ya ölçüm eş varyanslılığı nedir ve nasıl test edilir?

Ölçüm eş varyanslılığı (measurement invariance), iki veya daha fazla sayıda farklı kültür veya grup kapsamında yapılan ölçüm neticesinde kültürler veya gruplararası yapı varyansının (trait variance) farklı olmaması durumudur (Steenkamp ve Baumgartner, 1998). Daha basit bir ifadeyle, eğer Türkiye ve Malezya kültürlerinde yapılan materyalizm ölçümü eş varyans (invariance) ise bu iki farklı kültürden elde edilen verinin aynı yapıyı temsil eder niteliğe sahip olduğu çıkarılır. Böylece bu iki kültür arası materyalizm ortalaması üzerinden yapılacak karşılaştırma, Türk ve Malezyalı cevaplayıcıların materyalist eğilimlerinin karşılaştırılması olacaktır. Peki ya ölçüm eş varyanslılığı (measurement invariance) testi nasıl yapılmalıdır?

Ölçüm eş varyanslılığı testi üç aşamalı olarak gerçekleştirilmektedir (Steenkamp ve Baumgartner, 1998; Vandenberg ve Lance, 2000). İlk aşama olan yapısal eş varyanslılık (configural invariance) testinde ölçüm modeli (measurement model) tüm kültürler için eş anlı olarak kovaryans temelli yapısal eş itlik modellemesi kullanılarak analiz edilir. Yapısal eş varyanslılık testi, ölçüm modelinin tüm kültürlerde aynı şekilde gerçekleşip gerçekleşmediğ inin testidir. Daha basit bir ifadeyle, gizil yapının boyutsal yapısının ve ifadelerin dağılımının kültürlerarası istatistiksel olarak farklı olmadığ ının testi yapısal eş

varyanslılık testidir. Eğer yapısal eşvaryanslılık sağlanmış ise bir sonraki aşama olan metrik eşvaryanslılık (metric invariance) testine geçilir. Metrik eşvaryanslılık testinde ise gizil yapının altında yer alan boyutları oluşturan ifadelerin yüklenimlerinin (loadings) kültürlerarası farklılaşım farklılaşmadığı test edilmektedir. Bir diğer deyişle, ifadelerin ait oldukları boyutlara veya gizil yapıya yüklenimleri kültürlerarası istatistiksel olarak farklı değil ise tam metrik eşvaryanslılık (full metric invariance) sağlanmış olur. Eğer tam metrik eşvaryanslılık sağlanamamış ise her bir boyut altında en az bir ifadenin yükleniminin kültürlerarası istatistiksel olarak farklılaşmadığı ortaya konulursa kısmi metrik eşvaryanslılık (partial metric invariance) sağlanmış olur (Byrne vd., 1989). Metrik eşvaryanslılık testini uygularken yapısal eşitlik modellemesini kapsamında ifadelerden boyutlara giden yüklenim yolları (loading paths) gruplararası eş kılınır. Kısmi veya tam metrik eşvaryanslılık sağlandığında bir sonraki aşama olan sayısal eşvaryanslılık (scalar invariance) testine geçilir. Sayısal eşvaryanslılık testinde gizil yapının altında yer alan boyutları oluşturan ifadelerin yüklenimlerinin (loadings) yanısıra modeldeki tüm sabit terimlerinin (intercepts) de gruplararası eş kılındığı model analiz edilir. Eğer gruplararası gerçekleştirilen eş anlamlı uygulama sonucu istatistiksel olarak anlamsız tespit edilirse, yani, faktör yüklenimleri ile sabit terimler gruplararası istatistiksel olarak farklı değilse tam sayısal eşvaryanslılık sağlanmış demektir. Tam sayısal eşvaryanslılığın sağlanmadığı durumlarda her bir boyut altında en az bir faktör yüklenimi ile en az bir sabit teriminin kültürlerarası istatistiksel olarak farklılaşmadığı ortaya konulursa kısmi sayısal eşvaryanslılık (partial scalar invariance) sağlanmış olur (Byrne vd., 1989).

Yukarıdaki paragrafta değinilen üç tür eşvaryanslılık da sağlandığında ölçüm eşvaryanslılığın ilgili kültürlerarası var olduğu ve farklı kültürlerden elde edilen verinin gizil yapıya ilişkin yapılacak karşılaştırmalara uygun olduğu çıkarılabilir. Ölçüm eşvaryanslılığı testinin ilk aşaması olan yapısal eşvaryanslılık (configural invariance) sağlandığında araştırma şu çıkarımı sağlamaktadır: ‘Ölçümlemeye çalıştığım gizil yapı, analizini yaptığım iki veya daha fazla kültürde aynı

şeyi ifade ediyor'. Fakat araştırmacının amacı gizil yapıya ilişkin iki veya daha fazla kültürden toplanan verilerdeki ortalamalarını karşılaştırmak ise en azından kısmi metrik eşvaryanslılık (partial metric invariance) ile kısmi sayısal eşvaryanslılık (partial scalar invariance) sağlanmış olmalıdır. Araştırmacının amacı gizil yapının model içinde yer alan diğer gizil yapılar ile ilişkilerini kültürlerarası karşılaştırmak ise tam metrik eşvaryanslılık (full metric invariance) veya kısmi metrik eşvaryanslılık (partial metric invariance) sağlamış olmak yeterli olup sayısal eşvaryanslılık (scalar invariance) sağlamış olmaya gerek yoktur (Steenkamp ve Baumgartner, 1998).

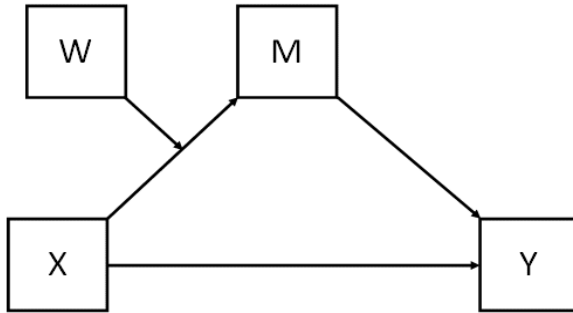
Toparlamak gerekirse, iki veya daha fazla sayıdaki kültürden psikometrik ölçekler aracılığıyla toplanmış olan verinin karşılaştırmasını yapmadan önce ölçüm eşvaryanslılığı (measurement invariance) sağlanmış olunmalıdır. Aksi takdirde, elde edilecek olan farklılık testi sonuçlarının gerçekten karşılaştırılması yapılan gizil yapı kaynaklı mı yoksa kullanılan yöntem veya hata varyansı kaynaklı mı olduğu bilinemez. Nitekim davranışsal ve sosyal bilimlerdeki araştırmacıların kültürlerarası verileri analiz ederken temel amaçları ortalama karşılaştırması yapmak değil gizil yapının ortalamalarını karşılaştırmaktır.

(4) İlimlaştırmacı (moderator) değişken ile aracı (mediator) değişkenin ardında yatan mantık nedir?

Davranışsal ve sosyal bilimlerdeki araştırmacılar, araştırma modelleri kurmakta ve bu modellerin içinde yer alan değişkenler (gizil yapılar) arası ilişkileri test etmektedirler. Değişkenler arası ilişkileri test etmenin ötesinde bu ilişkilerin ardında yatan mekanizmaların ve ilgili ilişkilerin koşullara bağlı olarak nasıl değiştiğinin keşfedilmesi de bilimsel bilgiyi artırmanın bir yoludur. Bir diğer deyişle, değişkenler arası ilişkilerin tespiti temel bir bilimsel bulgu iken bu ilişkilerin hangi koşullarda nasıl değiştiği ve bu ilişkilerin ardında hangi mekanizmaların bulunduğu açığa çıkarılması daha ileri düzeyde bir bilimsel bulgu ortaya koymayı ifade etmektedir. Bu ileri düzey bilimsel bilgi gelişimi sağlayabilmek adına ilimlaştırmacı (moderator)

değişken ile aracı değişken (mediator) kavramları hakkında bilgi sahibi olmak gereklidir. Peki ya ılımlatırıcı değişken ile aracı değişken araştırmacılar için ne ifade ediyorlar?

Bağımsız değişken (X) ile bağımlı değişken (Y) arasındaki ilişkinin yönü ve kuvveti başka bir değişkenin (Mod) seviyesine bağlı olarak farklılaşıyor ise bu durumda ılımlatırıcı değişkenin varlığı söz konusudur (Hayes, 2013). Örneğin; aylık gelir ile lüks tüketim harcaması arasındaki ilişki cinsiyete bağlı olarak farklılaşıyor ise cinsiyet ılımlatırıcı değişken rolüne sahiptir çıkarımında bulunulabilir. Detaylandırmak gerekirse, erkek cevaplayıcılar kapsamında aylık gelir ile lüks tüketim harcaması ilişkili değil ($r = .00, p > .05$) ise fakat kadın cevaplayıcılar kapsamında aylık gelir ile lüks tüketim harcaması pozitif anlamlı ilişkili ($r = .50, p < .05$) ise cinsiyet değişkeninin aylık gelir ile lüks tüketim harcaması arasındaki ilişkiyi düzenlediği veya ılımlatırdığı anlaşılır. Daha önce ifade edildiği gibi bu noktada cinsiyet değişkeni ılımlatırıcı değişken rolündedir. Öte yandan, bağımsız değişken (X) ile bağımlı değişken (Y) arasındaki ilişki başka bir değişken (Med) üzerinden gerçekleşiyor ise bu durumda aracı değişkenin varlığı söz konusudur (Hayes, 2013). Örneğin; aylık gelir ile lüks tüketim harcaması arasındaki ilişki materyalizm üzerinden gerçekleşiyor ise materyalizm aracı değişken rolüne sahiptir çıkarımında bulunulabilir.



Şekil 1: ılımlatırıcı değişken (W) ve aracı değişken (M)

Son 30 yıldır davranışsal ve sosyal bilimler kapsamında ılımlaştırıcı değişken ve aracı değişken kavramları ilgi görmektedir (Baron ve Kenny, 1986). Özellikle, ılımlaştırıcı etki ve aracı etkinin nasıl tespit edileceğine veya nasıl analiz edileceğine ilişkin tartışmaların yaygınca yapıldığı ve genel kabul görmüş uygulamaların literatürde yer aldığı bilinmektedir (Baron ve Kenny, 1986; Zhao vd., 2010; Darlington ve Hayes, 2017). Bu teknik tartışmalar, ılımlaştırıcı etki ve aracı etkinin ardında yatan felsefenin anlaşılmasına yönelik tartışmaları gölgede bırakmıştır. Ancak literatürde yer alan ılımlaştırıcı etki ve aracı etkiye ilişkin istatistiksel uygulamalardaki çarpıklıklar, araştırmacılar tarafından ılımlaştırıcı değişken ile aracı değişkenin felsefesinin anlaşılması hususunda sorunlar olduğuna işaret etmektedir (Aguinis vd., 2016). Peki ya ılımlaştırıcı değişken ile aracı değişkenin ardında yatan felsefe nedir?

En basit haliyle, ılımlaştırıcı değişken ilişkilere ilişkin ne zaman? sorusuna cevap sunarken aracı değişken ise ilişkilere ilişkin neden? sorusuna cevap sunar. Örneğin; aylık gelir ile lüks tüketim harcaması arasında kadınlar kapsamında anlamlı bir ilişki var iken erkekler kapsamında ise bir ilişki olmadığını varsayalım. Bu durumda ılımlaştırıcı değişken olan cinsiyet, ne zaman aylık gelir ile lüks tüketim harcaması ilişkilidir? sorusuna cevap sunmaktadır. Öte yandan aylık gelir ile lüks tüketim harcaması arasındaki ilişki materyalizm üzerinden gerçekleşiyor olsun. Bu durumda aracı değişken olan materyalizm, neden aylık gelir ile lüks tüketim harcaması ilişkilidir? sorusuna cevap sunmaktadır.

Teori geliştirme sürecinde ılımlaştırıcı değişken(ler) ile aracı değişken(ler)in rolü belirleyici niteliktedir. Çünkü teorileştirme süreci neyin (what), ne zaman (when) ve neden (why) gerçekleştiğine cevap sunmanın gerekli olduğu bir süreçtir (Sutton ve Staw, 1995; Weick, 1995). Bir diğer deyişle, teoriler bize neyin, ne zaman ve neden gerçekleştiğini sistematik şekilde anlatan hikayelerdir. Sosyal psikoloji literatüründen benlik kontrol gücü teorisi (self-control strength teorisi: Baumeister vd., 2007) örneği üzerinden tartışmak gerekirse, bu teori benlik kontrolüne ilişkin bir teoridir. Yani bireylerin kendi iç-

sel dürtülerine karşı kendilerini nasıl kontrol altında tuttukları, bu teorinin ne (what) sorusuna verdiği cevaptır. Ayrıca benlik kontrol gücü teorisi, insanların sınırlı bir benlik kontrol kapasitesi olduğunu önererek her bir benlik kontrolü girişimi sonrasında benlik kontrol kapasitesinde yaşanacak azalmadan dolayı sonraki benlik kontrolü gerektiren durumda bireyin kendini kontrol etmede başarısız olacağını ileri sürmektedir. Dahası, benlik kontrol kapasitesinin şeker alımı ile doldurulabileceği bu teori tarafından ileri sürülmektedir. Böylece insanların kendilerini kontrol etme sürecinde neden (why) başarısız olduklarına ilişkin de ilgili teori açıklama getirmektedir. Teorinin savlarının özellikle sınırlı bir iradeye sahip olduğunu düşünen insanlar kapsamında geçerli olduğu fakat sınırsız bir iradeye sahip olduğunu düşünen insanlar kapsamında geçerli olmadığı da yakın geçmişte ulaşılan bir bulgudur (Job vd., 2013). Bu bulgu, teorinin ne zaman çalıştığına dair cevap sunmaktadır. Böylece benlik kontrol gücü teorisi ne (what), ne zaman (when) ve neden (why) sorularına sistematik bir cevap sunmaktadır.

Benlik kontrol gücü teorisi (Baumeister vd., 2007) örneği üzerinden resmedilen teorinin bileşenleri ve teorileştirme süreci anlaşılabilir olduğu üzere soyut ve fiktisel bir süreçtir. Teorileştirme yapılırken yapılan önermelerin test edilebilmesi adına ılımlaştırıcı (moderator) değişken ve aracı (mediator) değişken kavramları hizmet etmektedirler. Bu perspektiften, ılımlaştırıcı (moderator) değişken ve aracı (mediator) değişkenin ne zaman ve neden sorularına cevap vererek teorileştirme sürecine hizmet ettiği görülmektedir. Dolayısıyla davranışsal ve sosyal bilimlerdeki araştırmacıların, ılımlaştırıcı (moderator) değişken ve aracı (mediator) değişkeni sadece birer istatistiksel terim olarak kavraması yerine teorileştirme sürecinin bir parçası olarak kavraması önerilmektedir.

(5) Neden etkileşim terimi içeren bazı regresyon analizi sonuçları sorunludur ve neden ortalama merkezleme (mean-centering) işlemi yapılmalıdır?

Regresyon analizi, bağımsız değişken(ler) aracılığıyla bağımlı değişkeni tahmin etme (kestirme) imkanı sağlayan bir istatistiksel tek-

niktir (Cohen vd., 2003). Basit doğrusal regresyon modelinde (simple linear regression model) bir adet bağımsız değişken ile bağımlı değişken tahmin edilmeye çalışılırken, çoklu doğrusal regresyon modelinde (multiple linear regression model) ise birden fazla sayıda bağımsız değişken ile bağımlı değişken tahmin edilmeye çalışılmaktadır. Basit doğrusal regresyon modeli matematiksel olarak $Y = \beta_0 + \beta_1 X_1$ şeklinde ifade edilebilir. Bu denklemde yer alan Y ifadesi bağımlı değişkeni, X_1 ifadesi bağımsız değişkeni, β_0 sabit terimi ve β_1 ise eğim (slope) olarak da bilinen regresyon katsayısı ifade etmektedir. Çoklu doğrusal regresyon modeli (iki bağımsız değişkene ve bir sabit terime sahip) matematiksel olarak $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ şeklinde ifade edilebilir.

Basit doğrusal regresyon modeline ilişkin analiz sonuçlarını yorumlamak çoklu doğrusal regresyon modeline ilişkin analiz sonuçlarını yorumlamaya kıyasla daha basittir. Basit doğrusal regresyon modelinde yer alan β_1 katsayısı, X_1 bağımsız değişkenindeki bir birimlik artış neticesinde Y bağımlı değişkenindeki bir birimlik artışın ne kadar olacağını ifade etmektedir. Benzer mantıkla çoklu doğrusal regresyon modelini yorumlamak mümkün olmakla birlikte detaylara dikkat etmek kritik öneme sahiptir. Yukarıda yer alan çoklu doğrusal regresyon modeli kapsamında β_1 katsayısı (basit doğrusal regresyon modelinden farklı olarak), X_2 bağımsız değişkeni sabit tutulduğunda X_1 bağımsız değişkenindeki bir birimlik artış neticesinde Y bağımlı değişkenindeki bir birimlik artışın ne kadar olacağını ifade etmektedir. Peki ya X_2 bağımsız değişkeni nasıl sabit tutulabilmektedir? Burada X_1 bağımsız değişkenindeki bir birimlik artış neticesinde Y bağımlı değişkenindeki bir birimlik artışın X_2 bağımsız değişkeninin tüm değerleri için hesaplaması yapılır ve ortalaması alınır. Elde edilen ortalama değer bu noktada β_1 katsayısına eşittir. Bir diğer deyişle, X_2 bağımsız değişkeninin tüm değerleri için β_1 katsayısı hesaplanıp ortalaması alındığında X_2 değişkeninden bağımsız bir β_1 katsayısı hesaplanmış olur. Bu durum, şartsız regresyon modeli (unconditional regression model; Darlington ve Hayes, 2017) olarak da isimlendirilmektedir. Peki ya çoklu doğrusal regresyon modeline etkileşim terimi eklenirse, sonuçlar aynı şekilde yorumlanabilir mi?

Etkileşim terimi, iki yada daha fazla sayıdaki değişkenin çarpımı neticesinde elde edilen bir terimdir (Cohen vd., 2003). Örneğin; $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$ modelinde yer alan X_1X_2 değişkeni etkileşim terimidir. X_1 ve X_2 bağımsız değişkenlerinin çarpımı ile elde edilen etkileşim terimi özellikle ılımlaştırıcı etki (moderator) analizinde kullanılmaktadır. Yukarıda yer alan β_3 katsayısı istatistiksel olarak anlamlı tespit edildiğinde ılımlaştırıcı etkisi olduğu çıkarımı sağlanmaktadır (Baron ve Kenny, 1986). Ancak etkileşim teriminin regresyon modeline dahil edilmesinin β_1 ve β_2 katsayılarının yorumlanmasında bir sorunu beraberinde getirmektedir. Bu sorun, davranışsal ve sosyal bilimlerde sıklıkla analiz(ler) konu olan psikometrik ölçek(ler) aracılığıyla toplanmış verileri kapsamaktadır. Detaylandırmak gerekirse, etkileşim teriminin regresyon modeline dahil edilmiş olması şartsız (unconditional) regresyon modelini şartlı (conditional) regresyon modeline dönüştürmektedir. Teknik olarak ifade etmek gerekirse; etkileşim terimi modele dahil edildiğinde, β_1 katsayısı X_2 bağımsız değişkeni sabit tutulduğunda X_1 bağımsız değişkenindeki bir birimlik artış neticesinde Y bağımlı değişkenindeki artışı ifade etmekte olup X_2 bağımsız değişkeni sıfır'a eşit olduğunda ($X_2 = 0$) X_1 bağımsız değişkenindeki bir birimlik artış neticesinde Y bağımlı değişkenindeki artışı ifade etmektedir. Bir anlamda β_1 katsayısı hesabı $X_2 = 0$ durumuna koşullanmaktadır. Aynı şekilde β_2 katsayısı hesabı $X_1 = 0$ durumuna koşullanmaktadır. Bu nedenle şartlı regresyon modeline dönüşüm gerçekleşmektedir (Hayes, 2013). Böyle bir durumda, X_1 veya X_2 değişkeninin psikometrik ölçüm araçları aracılığıyla ölçümlendiği (5'li likert vb.) ve X_1 ya da X_2 bağımsız değişkeninin sifıra eşit olmasının mümkün olmadığı için β_1 ve β_2 katsayılarına yönelik yorumlamalar geçersiz olacaktır. Ayrıca β_1 ve β_2 katsayılarının istatistiksel olarak anlamlılığına ilişkin test sonuçları da geçersiz olacaktır (Aiken ve West, 1991; Darlington ve Hayes, 2017). Bu sorunun çözümü ortalama-merkezleme (mean-centering) işlemi yaptıktan sonra etkileşim terimini oluşturmaktır. Bir diğer deyişle, ortalama merkezleme işlemi yapıldıktan sonra bağımsız değişkenler çarpımı yapılarak etkileşim terimi oluşturulmalıdır. Ortalama-merkezleme iş-

lemi ise her bir değişkene ait ham veriden o değişkenin ortalaması çıkartılarak yapılmaktadır (Hayes, 2013). Ortalama-merkezleme işlemi sonrasında oluşturulan etkileşim terimini içeren $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_1X_2$ regresyon denkleminin analizinde β_1 katsayısı X_2 'in sifıra eşit olduğu koşul için değil de X_2 'in kendi ortalamasına eşit olduğu koşul için hesaplanacaktır. Böylece ortalama değer ölçüm aralığı içinde yer alan gerçek bir değer olacağı için (5'li likert ölçüm neticesinde 3.25 ortalama gibi) β_1 ve β_2 katsayılarının yorumlanmasında sorun olmayacaktır.

Davranışsal ve sosyal bilimlerdeki araştırmacılar özellikle ılımlaştırıcı etki testi yapabilmek adına etkileşim terimini regresyon modeline dahil ettiklerinde bu başlık altında yapılan uyarıyı göz önünde bulundurmalarıdır. Aksi takdirde, bağımsız değişken(ler)in bağımlı değişkeni istatistiksel anlamlı şekilde kestirebilip kestiremediği veya ne düzeyde kestirdiğine yönelik çıkarımları gerçek dışı bir varsayım üzerine kurulu olacaktır. Ancak bağımsız değişkenlerin sıfır değerini alabildiği durumda (gelir vb.) böyle bir tehdit söz konusu değildir.

(6) p-değeri nedir?

Davranışsal ve sosyal bilimlerdeki araştırmacılar hipotezler geliştirip bu hipotezleri test ederek bilimsel bilgi birikimi oluşturabilmektedirler. Bu alandaki araştırmacılar yaygın şekilde boş (null) hipotez ile alternatif hipotezi belirleyerek hipotez test etme sürecine başlamaktadırlar. Ancak hipotez test sürecine ilişkin yakın geçmişte davranışsal ve sosyal bilimlerdeki araştırmacıların p-değerine ilişkin uygulamalarının bilinçsizce gerçekleştiği dikkat çekmektedir (Gigerenzer, 2004). Detaylandırmak gerekirse; hipotez test etme süreci, Fisher (1955) tarafından önerilen boş hipotez testi (null hypothesis testing) yaklaşımı doğrultusunda veya Neyman (1950) karar teorisi (decision theory) olarak bilinen yaklaşım doğrultusunda gerçekleştirilebilir. Fisher (1955)'ın boş hipotez testi yaklaşımına göre; (1) boş hipotez belirlenir, (2) istatistiksel anlamlılık seviyesi belirlenir, (3) veri toplanıp test edilir ve sonuç olarak boş hipotezin red edilebilip edilemediği yönünde çıkarım sağlanır. Bu yaklaşım özellikle hipotezin keşifsel

doğaya sahip olduğu ve test edilecek net bir önermenin olmadığı durumda uygulanması idealdir. Görüldüğü üzere Fisher (1955)'ın boş hipotez testi yaklaşımında alternatif hipotez belirlenmemektedir. Öte yandan Neyman (1950) karar teorisi yaklaşımına göre; (1) iki farklı hipotez olan H1 ve H2 ile tip-1 hata payı (α), tip-2 hata payı (β) ve örneklem büyüklüğü belirlenir, (2) veri toplanıp analiz edildikten sonra elde edilen test değeri H1 hipotezin reddini öneriyorsa H2 kabul edilir veya test değeri H2 hipotezin reddini öneriyorsa H1 kabul edilir.

Fisher (1955)'ın hipotez testine yaklaşımı boş hipotez (H0) üzerine odaklanmış iken Neyman (1950)'ın yaklaşımı ise iki farklı alternatif hipotez olan H1 ve H2 kapsamında ikili bir seçime odaklanmaktadır. Peki ya günümüzde davranışsal ve sosyal bilimlerde uygulanan hipotez testleri hangi yaklaşımı izliyor? Bu sorunun cevabını Gigerenzer (2004) iki farklı hipotez testi yaklaşımının tutarsız ve çelişkili bir karışımı olarak veriyor. Bir diğer deyişle, iki farklı hipotez testi yaklaşımı kombine edilerek boş hipotez anlamlılık testi (null hypothesis significance testing; Nickerson, 2000) olarak bilinen hipotez testi yaklaşımı davranışsal ve sosyal bilimlerde yaygın şekilde uygulanmaktadır. Boş hipotez anlamlılık testine göre; (1) sadece boş hipotez belirlenir ve (2) istatistiksel anlamlılık testi sonucuna göre ya boş hipotez ret edilir ya da boş hipotez ret edilemez. Görüldüğü üzere boş hipotez anlamlılık testinin ilk adımı Fisher (1955)'ın tavsiyesini takip ederken ikinci adımı ise Neyman (1950)'ın tavsiyesini takip etmektedir. Bu tutarsızlık, davranışsal ve sosyal bilimlerde hipotez testi sürecinin net bir şekilde kurgulanamamış olmasını gözler önüne sermektedir. Hatta Meehl (1978) boş hipotez anlamlılık testini psikolojinin başına gelmiş en kötü şey olarak nitelendirmektedir.

Davranışsal ve sosyal bilimlerdeki araştırmacılar tarafından benimsenen boş hipotez anlamlılık testi sürecinde en kritik rol p-değerine atfedilmiştir. Çünkü belirlenen güven aralığı ve tip-1 hataya bağlı olarak (örn: $p < .05$) belirli bir kesim değerinin altında yer alan p-değerine sahip bulgular istatistiksel olarak anlamlı şekilde yorumlanırken bu kesim değerinin çok az da olsa üzerinde p-değerine sahip

bulgular istatistiksel olarak anlamsız şekilde yorumlanmaktadır. Örneğin; $p = .049$ değerine sahip bir hipotez testi sonucu anlamlı iken $p = .051$ değerine sahip bir hipotez testi sonucu anlamsızdır. Üstelik davranışsal ve sosyal bilimler kapsamında yer alan akademik dergilerin sadece anlamlı hipotez testi sonuçlarını yayınladıkları göz önünde bulundurulduğunda, adeta $p = .050$ değeri bir ölüm-kalım çizgisi niteliğine sahip olmaktadır. Bu sebepten dolayı veriyle oynama (data fabrication) gibi etik olmayan araştırma eylemlerinin gerçekleşmesinde p -değerine bağlı olarak anlamlılık testlerinin varlığı suçlanmaktadır (Cumming, 2014). Peki ya hipotez testlerinde bu derece önem atfedilen p -değeri ne anlama gelmektedir?

Düşünün ki yaptığınız hipotez testi neticesinde p -değerini .02 olarak tespit ettiniz. Bu bulgunuz tam olarak şunu ifade ediyor: alternatif hipotezin (H1) doğru olduğu varsayımı altında hipotez testini defalarca tekrarlıyorsanız yapmış olduğunuz hipotez testlerinin %2'sinde H1 hipotezini yanlışlıkla red etmiş olacaksınız (Gigerenzer, 2004). Buradaki açıklamadan anlaşılabilceği üzere, p -değeri bize alternatif hipotezimizin (H1) ne derece doğru olduğu hakkında birşey söylememektedir. Çünkü p -değerine ilişkin hesaplamalarda alternatif hipoteze ilişkin doğru olduğu yönünde bir varsayım yapılmaktadır. Dolayısıyla varsayım doğru veya yanlış olabilir. Fakat araştırmacı p -değeri üzerinden olasılıksal anlamda alternatif hipotezin doğruluğuna ilişkin bir çıkarım sağlayamaz. Benzer şekilde p -değeri bize başka bir örneklemden veri topladığımızda aynı sonuca ulaşma olasılığımızı da söylememektedir. Bir diğer deyişle, Eskişehir'deki 250 üniversite öğrencisinden toplanan verinin analizi neticesinde X ile Y değişkeni arasında pozitif anlamlı bir ilişki ($r = .50$, $p = .03$) tespit edilmiş ise aynı çalışmayı Karaman'daki 250 üniversite öğrencisinden veri toplayarak tekrarlamak istediğimizde elde edeceğimiz sonuca ilişkin $p = .03$ değeri hiçbir çıkarım sağlamamaktadır. Bu yüzden p -değeri oldukça düşük bulgulara daha fazla güvenmek gibi bir içgörü sağlamak mümkün değildir.

p -değeri, hipotez testi sonuçlarının ne düzeyde şans eseri elde edildiğine dair bir çıkarım sağlayabilmektedir (Nuzzo, 2014). Ancak

bu noktada p-değerinin ana varsayımı olan alternatif hipotezin doğru olduğu varsayımı akıllardan çıkartılmamalıdır. Araştırmacı tarafından geliştirilen ve ileri sürülen alternatif hipotez ne düzeyde gerçektir? Örneğin, 'dindarlık ile kullanılan otomobil markası arasında anlamlı bir ilişki vardır' ve 'dindarlık ile gönüllü sadelik arasında anlamlı bir ilişki vardır' hipotezleri alternatif hipotezlerdir. Fakat ilk hipotez, ikinci hipoteze kıyasla daha az inanılır niteliğe sahiptir. Her ne kadar alternatif hipotezlerimizi teoriler ve düşüncel tartışmalar doğrultusunda geliştirebiliyor olsak da geliştirilen alternatif hipotezin doğru olduğunu varsaymak neredeyse imkansızdır. Bu yüzden $p = .03$ tespit edildiğinde, %3 olasılıkla elde edilen sonuç şans eseri gerçekleşmiştir yorumu yapılırken alternatif hipotezin doğru olduğu durumda ilgili yorumun geçerli olduğu unutulmamalıdır. P-değerinin bize ifade edeceği anlam alternatif hipotezin doğru olduğu koşula dayandığından dolayı alternatif hipotezlerin ne derece güçlü şekilde geliştirilmiş olduğu kritik öneme sahiptir. Ne kadar güçlü şekilde teorik arkaplana ve gerekçeye sahip alternatif hipotezlere sahip isek o düzeyde p-değerlerinin anlam ifade ettiği bir literatüre sahip olabiliriz.

(7) Hipotez testi sonuçları nasıl yorumlanmalıdır? Neden boş hipotez (H0) kabul edilemez?

Davranışsal ve sosyal bilimlerdeki araştırmacılar yaygın olarak hipotez test etme sürecinde boş hipotez anlamlılık testi (null hypothesis significance testing) yaklaşımını izlemektedirler. Detaylandırmak gerekirse, boş (null) ve alternatif hipotezler kurgulanıp istatistiksel anlamlılık seviyesini belirleyici detaylar (tip-1 hata, tip-2 hata) belirlendikten sonra verinin analizi gerçekleştirilir. Analiz neticesinde elde edilen test istatistiği değerine (p-değeri) göre çıkarım sağlanmaktadır. Bu çıkarımlar rapor edilirken, ulusal araştırmacılar arasında yaygın şekilde alternatif hipotezin (H1) kabul edildiği yönünde vurgular kullanıldığı bilinmektedir. Peki ya boş hipotez anlamlılık testi neticesinde alternatif hipoteze ilişkin böylesine net bir çıkarım elde etmek mümkün müdür?

Boş hipotez anlamlılık testi (null hypothesis significance testing) sürecinin merkezinde boş hipotez yer almakta olup tüm istatistik-

sel hesaplamalar boş hipoteze ilişkin çıkarım sağlamaktadır. Ancak boş hipotezin tam zıttında yer alan alternatif hipoteze ilişkin de bazı çıkarımları boş hipotez anlamlılık testi neticesinde çıkarmak mümkündür. Detaylandırmak gerekirse, hipotez testi neticesinde $p = .15$ (%95 güven aralığı) değerine ulaşan bir araştırmacı boş hipotezi red edememektedir. Bir diğer deyişle, 'boş hipotez red edilememiştir' şeklinde bir çıkarım sağlanmıştır. Ayrıca 'alternatif hipotez red edilmiştir' çıkarımı sağlanamaz. Bu noktada boş hipotezin red edilememiş olması alternatif hipotezin red edildiği veya kabul edilmediği anlamına gelmemektedir (Altman ve Bland, 1995). Öte yandan, hipotez testi neticesinde $p = .01$ (%95 güven aralığı) değerini tespit eden bir araştırmacı ise boş hipotezi red edebilmiştir. Böyle bir durumda araştırmacı 'alternatif hipotez kabul edilmiştir' sonucuna ulaşabilir.

Neden alternatif hipotezin red veya kabul edildiği yönünde hipotez sonuç vurgusunda bulunmanın sakıncalı olduğunu bilim felsefesi temelli olmak üzere açıklamak mümkündür. Popper (1959)'ın yanlışlamacı yaklaşımına göre bilimsel bilgi birikimi birikimsel olarak değil yanlışlayarak ilerlemektedir. Popper (1959)'ın meşhur siyah kuğu metaforunda yer aldığı üzere, 'evrendeki tüm kuğular beyazdır' boş hipotezimizi test etmek amacıyla kuğuları gözlemlemeye başladığımızı düşünelim. Yaşadığımız şehirdeki kuğuları gözlemledikten ve gözlemlediğimiz kuğuların hepsinin beyaz olduğunu tespit ettikten sonra 'evrendeki tüm kuğular beyazdır' boş hipotezimizi kabul ettiğimizi varsayalım. Yıllar sonra farklı bir ülkedeki parkta siyah kuğu gördüğümüzde daha önce kabul etmiş olduğumuz 'evrendeki tüm kuğular beyazdır' boş hipotezi ortadan kalkacaktır. Bir diğer deyişle, birikimsel şekilde ilerlemeden ziyade yanlışlayarak bilimsel bilgi birikimi gelişmektedir. Dolayısıyla çelişkileri ortadan kaldıracı bir hipotez red edilerek veya red edilemeyerek hipotez testleri sonuçları bilimsel bilgi birikimini ilerletmektedir (Dienes, 2008).

Toparlamak gerekirse, boş hipotez anlamlılık testi sürecinde hipotez test sonuçları yorumlanırken boş hipotezin kabul edilmesi gibi bir çıkarımda bulunmak söz konusu değildir. Boş hipotez (H_0) red

edilebilir ki bu durum alternatif hipotezin (H1) kabul edildiği çıkarımına sunar. Boş hipotezi (H0) red edilemeyebilir ki bu durum alternatif hipotezin (H1) red edildiği çıkarımını sağlamaz. Çünkü ilerleyen süreçte boş hipotezi red edecek bir gözlem ile karşılaşılabılır ve böylece alternatif hipotez kabul edilebilir.

(8) Neden veriye işkence edilmemelidir?

Bilimsel araştırma projeleri kapsamında toplanmış olan veri, birçok araştırmacı tarafından konuşturulmayı bekleyen bir kaynak olarak yorumlanabilir. Bu yaklaşım ‘veriye işkence edilirse veri konuşur’ anlayışının gelişmesine yardımcı olmaktadır. ‘Veriye işkence etme’ fikri beraberinde fırsatları ve tehlikeleri sunan bir yaklaşımdır. Yakın geçmişte sosyal psikoloji alanında yaşanan replikasyon krizi (replication crisis; Open Science Collaboration, 2015) ile birlikte birçok metodolog ‘veriye işkence etme fikrinin’ bir sonucu olarak bu krizi yorumlamaktadırlar (Simmons vd., 2011; Simonsohn, 2015; Munafò vd., 2017). Peki ya ‘veriye işkence etme’ ne demek? ve ‘veriye işkence etme’ fikri ne tür fırsatları ve tehditleri beraberinde getirmektedir?

İlk olarak veriye işkence etmenin kavramsal çerçevesini çizmek adına ‘veriye işkence etme’ ile ‘veriyi analize hazırlama’ ayrımı yapılacaktır. Davranışsal ve sosyal bilimlerde yaygın olarak gerçekleştirilen psikometrik ölçüm araçları aracılığıyla veri toplama süreci neticesinde araştırmacılar ham veriyi elde etmektedirler. Araştırma hipotezlerini test etmeye yönelik gerçekleştirilecek olan istatistiksel analizler öncesinde ham verinin ilgili analizlere hazırlanması önemli bir süreçtir. Verinin normal dağılım sergileyip sergilemediğinin kontrolü, ters kodlu ifadelerin gereken şekilde kodlanıp kodlanmadığının kontrolü, uç (outlier) verilerin veri setinden çıkarılması, psikometrik olarak beklenen şekilde hareket etmeyen ifadelerin veri setinden çıkarılması gibi kararlar ‘veriyi analize hazırlama’ işlemleri olarak kavramsallaştırılabilir. Ancak bu işlemler haricinde veri seti kapsamında keşifsel analizler yapmak ve bu doğrultuda veri toplama süreci öncesinde öngörülemez bulgulara ulaşmak üzere veriye farklı açılardan bakmak ‘veriye işkence etmek’ olarak kavramsallaştırılabilir. Örne-

ğın; X ile Y değişkenleri arasında pozitif ilişkinin var olduğu yönünde kurgulanan hipotezin testi için gerçekleştirilen veri toplama sonrasında araştırmacı uyguladığı analizler neticesinde X ile Y arasındaki ilişkinin 0.05 düzeyinde anlamsız olduğu ($r = .03$, $p > .05$) tespit etmiş olsun. Ayrıca veriye işkence etme süreci neticesinde Z değişkeninin yüksek değerleri kapsamında X ile Y pozitif anlamlı ilişkili iken Z değişkeninin düşük değerleri kapsamında X ile Y arasında negatif anlamlı ilişki olduğu tespit edilmiş olsun. Bu durumda araştırmacı uygulamış olduğu veriye işkence süreci neticesinde aslında Z değişkeninin seviyesine bağlı olarak X ile Y arasında ilişkinin şekillendiği yönünde hipotezini revize etmesi durumunda HARKING¹ (hypothesizing after results are known; Kerr, 1998) olarak isimlendirilen hatayı gerçekleştirmiş olacaktır. Bir diğer deyişle, veriye işkence neticesinde elde edilen sonuçların sanki veri toplanmadan önce tahmin edilmiş gibi hipotezin revize edilmesini ifade eden HARKING, veriye işkence etmenin bir sonucudur. Peki ya bu işkence bir keşif ile sonuçlanmış iken ortadaki sorun nedir?

Örneklemeden toplanan verilerin istatistiksel analizleri uygulanırken toplanan verinin doğasından kaynaklı bazı hatalı sonuçlara ulaşma olasılığı her zaman araştırmacı tarafından göz önünde bulundurulmalıdır. Nitekim Tip-1 hata (α) ve Tip-2 hata (β) olarak bilinen terimler bu hata paylarına ilişkin çıkarımlar sağlamaktadırlar (Cohen vd., 2003). Tip-1 hata (α) yanlış şekilde alternatif hipotezin (H1) desteklenmesini ifade ederken, Tip-2 hata (β) ise yanlış şekilde boş hipotezin (H0) reddedilmesini ifade etmektedir. Bu iki tip hata kaynaklı olmak üzere her istatistiksel anlamlılık testi hatalı olma riski altındadır. Bu yüzden toplanan verinin gerçekliği mükemmel şekilde yansıttığını varsayıp veri analizleri neticesinde ulaşılan sonuçları gerçekliğin resmi olarak kabul etmek sakıncalı olacaktır. Ayrıca her toplanan veri kendi karakteristiğine sahiptir. Örneğin; 1000 bireyden oluşan bir popülasyondan alınan 50 kişiden oluşan her bir farklı alt

1 HARKING, analizler öncesinde değil analizler yapıldıktan sonra hipotezlerin geliştirilmesi yönünde yapılan yanlış yaklaşımdır.

örneklem neticesinde elde edilecek olan bulgular farklı olabilir. Nitekim her bir örneklemin popülasyonu ne derece temsil ettiği standard hata (standard error) düzeyiyle ifade edilmektedir. Örneklemin ortalaması popülasyonun ortalamasını ne düzeyde iyi temsil ediyorsa o düzeyde standard hata düşük olacaktır. Standard hata, regresyon ve yapısal eşitlik modeli analizleri kapsamında regresyon katsayıları ve yol katsayılarının istatistiksel olarak anlamlı olup olmadığının testinde kritik role sahiptir. Regresyon katsayısı veya yol katsayısının kendine ait olan standard hata'ya bölümü neticesinde elde edilen değer belirlenmiş olan kritik değerden yüksek olması durumunda sonuç istatistiksel olarak anlamlı şeklinde isimlendirilirken aksi takdirde istatistiksel olarak anlamlı değil şeklinde isimlendirilmektedir (Aiken ve West, 1991). Bu noktada standard hata değeri ne kadar düşükse örneklemin ortalamasının popülasyonun ortalamasını o derece iyi şekilde temsil ettiği çıkarılır.

Yukarıda değinilen tip-1 hata, tip-2 hata ve standard hata kavramlarından anlaşılacağı üzere, örneklemden elde edilen verinin analizi neticesinde ulaşılan sonuçların araştırmacıyı yanıltabileceği anlaşılmaktadır. Dolayısıyla farklı açılardan veriye bakarak önceden tahmin edilmemiş olan bulguları sanki veri toplamadan önce tahmin edilmiş gibi göstererek araştırmacılar literatürün gelişimine önemli ölçüde zararlar verebilmektedirler. Çünkü elde edilen bulgu o araştırmacının verisine has bir bulgu olabilir ve bu doğrultuda ilgili bulgu replikasyon edilmek istendiğinde başarısız olunma olasılığı yüksek olacaktır. Araştırmacıların amacı örneklemelere has bulgulara ulaşmak değil evrensel gerçeklikleri keşfetmektir. Bu yüzden teori-leştirme sürecinin ve bu süreç neticesinde ileri sürülen araştırma hipotezlerinin önemi büyüktür. Teorik düşünsel süreçler neticesinde ileri sürülen hipotezlere bağlı kalınmalı ve veriye işkence edilerek elde edilen farklı sonuçlar neticesinde başlangıçtaki hipotezler revize edilmemelidir.

'Veriye işkence etme' sadece olumsuzluklar değil avantajlar da sunabilen bir yaklaşımdır. Bu avantajları elde edebilmek için 'keşfedici

arařtırma (exploratory research)' bařlıđı sınırları ierisinde kalmak gereklidir. Bir diđer deyiřle, arařtırma hipotezinin kurgulanmadıđı bir ncl veri toplama sreci neticesinde elde edilen veriye iřkence edilerek bazı ilgin bulgulara ulařılabilir. Ancak bu keřfedici arařtırma sonrasında, keřifsel bulguları test etmeye ynelik olarak arařtırma hipotezinin kurgulandıđı ve test edildiđi ikinci bir arařtırma iin veri toplanmalı ve veriye iřkence edilmeden analizler gerekleřtirilmelidir (Armstrong, 1970; Miguel vd., 2014).

'Veriye iřkence etme' kararı verme srecinde arařtırmacı kendi teorik okumaları ve zihinsel tartıřmaları neticesinde ileri srdđ arařtırma hipotezine mi? yoksa elindeki verinin ona sylediđi sonulara mı? gvenmeliyim sorusunu kendine yneltmelidir. Her ne kadar sosyal psikoloji alanında gerekleřen replikasyon krizi (Open Science Collaboration, 2015) ncesine kadar verinin sylendiđinin takip edilmesi gerektiđi sekin arařtırmacılar tarafından nerilse de (Bem, 2003; Baumeister, 2016) gelecekte yařanabilecek olası replikasyon krizlerinin nne geebilmek iin bu nerileri terk edip teori-merkezli yaklařımı veri-merkezli yaklařıma tercih etmeliyiz (Wagenmakers vd., 2011).

(9) Aıklayıcı faktr analizi sonuları ile dođrulayıcı faktr analizi sonuları neden tutarlı olmayabilir?

Aıklayıcı faktr analizi, ok sayıdaki gzlemlenebilir ifadeyi onların iliřkili olmasını sađlayan gizil yapılar (latent constructs) altında konumlanacak řekilde sınıflamaya ynelik bir keřifsel analizdir (Thompson, 2004). Bu tanımlamadan anlařılabileceđi zere, aıklayıcı faktr analizi, gzlemlenebilir ifadelerin arka planda gizil yapıları temsil ettiklerini varsaymaktadır. Bu varsayım, temel bileřenler analizi (principal component analysis) kapsamında geerli deđildir. Bir diđer deyiřle, temel bileřenler analizi sadece ok sayıda ifadeyi onları temsil edebilecek az sayıda ifadeye dnřtrmeye odaklanmaktadır (Costello ve Osborne, 2005). Bu yzden temel bileřenler analizi ile aıklayıcı faktr analizi aslında aynı řeyi ifade etmemektedir. Bu ayrımaya rađmen, aıklayıcı faktr analizi yapma niyetiyle birok arařtır-

macının temel bileşenler analizi uyguladığı yaygınca karşılaşılan bir durumdur. Dolayısıyla, temel bileşenler analizi uygulayan bir araştırmacı, bulgularını doğrulayıcı faktör analizi kapsamında test edemez. Çünkü temel bileşenler analizi gizil yapıların varlığını hesaba katmaz iken doğrulayıcı faktör analizi ise gizil yapıların varlığı üzerine kurulu bir analizdir. Bilindiği üzere, doğrulayıcı faktör analizi kapsamında, araştırmacının kurguladığı modelin veri tarafından desteklenip desteklenmediği kovaryans temelli yaklaşımla test edilmektedir (Bagozzi ve Yi, 2012). Bu noktada doğrulayıcı faktör analizi uygulamaları, kovaryans temelli yapısal eşitlik modeli analizi aracılığıyla gerçekleştirilmektedir.

Açıklayıcı faktör analizi gerçekleştirilirken araştırmacının zihninde ölçüm ifadelerinin hangi gizil yapı veya psikometrik ölçeğe ait olduğuna dair bir ön bilgi yoktur. Bu yüzden açıklayıcı faktör analizi keşfetme niteliğine sahiptir. Üstelik İngilizce karşılığı ‘exploratory factor analysis’ olan açıklayıcı faktör analizinin keşfedici faktör analizi olarak isimlendirmek de daha doğru olacaktır. Öte yandan, doğrulayıcı faktör analizinde ise araştırmacının zihninde hangi ölçüm ifadelerinin hangi gizil yapı veya psikometrik ölçeğe ait olduğuna dair bir ön bilgi bulunmaktadır (Costello ve Osborne, 2005).

Açıklayıcı faktör analizi gerçekleştirirken araştırmacılar sıklıkla farklı çıkarma (extraction) ve döndürme (rotation) yöntemlerinden faydalanabilmektedirler. Bu çıkarma yöntemleri; temel bileşenler analizi (principle component analysis), ağırlıklandırılmamış en küçük kareler (unweighted least squares method), genel en küçük kareler yöntemi (general least squares method), en büyük benzerlik yöntemi (maximum-likelihood method), temel eksen boyutlandırma (principle axis factoring), alfa yöntemi (alpha method) ve imaj faktörlendirme (image factoring) şeklindedir. Her ne kadar SPSS programı içinde temel bileşenler analizi (principle component analysis) uygulaması açıklayıcı faktör analizi çıkarma seçeneği olarak konumlandırılrsa da önceki paragrafta tartışıldığı üzere temel bileşenler analizi aslında açıklayıcı faktör analizi ile aynı varsayıma sahip değildir. Do-

layısıyla açıklayıcı faktör analizi uygulama amacıyla temel bileşenler analizi çıkarma yönteminin seçilmesi önerilmemektedir. Öte yandan açıklayıcı faktör analizi kapsamında döndürme yöntemleri; varimax method, direct-oblimin method, quartimax method, equamax method ve promax rotation şeklindedir. Bu döndürme yöntemleri, ortogonal (orthogonal) yöntemler ve eğik yöntemler (oblique) yöntemler başlığı altında toplanmaktadır. Ortogonal döndürme yöntemlerinde (varimax, quartimax, equamax) faktörler arası ilişki olmadığı varsayılırken, eğik döndürme yöntemlerinde (direct-oblimin, promax) ise faktörlerin ilişkili olabileceği varsayılmaktadır (Thompson, 2004). Bu yüzden araştırmacı eđer faktörlerin ilişkili olabileceğini düşünüyorsa direct-oblimin veya promax yöntemi ile döndürme işlemini yapması önerilirken, eđer faktörlerin ilişkisiz olacağı öngörülüyorsa varimax, quartimax veya equamax yöntemi ile döndürme işlemini yapması önerilmektedir. Peki ya hangi çıkarma (extraction) ve döndürme (rotation) yaklaşımı tercih edilerek açıklayıcı faktör analizi uygulanmalıdır ki sonrasında uygulanacak doğrulayıcı faktör analizi ile tutarlı bir yol izlenmiş olsun?

Doğrulayıcı faktör analizi, kovaryans temelli yapısal eşitlik modeli aracılığıyla uygulandığından en büyük benzerlik yöntemi (maximum-likelihood method) kapsamında doğrulayıcı faktör analizi gerçekleştirilmektedir. Bu yüzden açıklayıcı faktör analizi uygularken çıkarma (extraction) yöntemi olarak en büyük benzerlik yönteminin (maximum-likelihood method) tercih edilmesi doğrulayıcı faktör analizi ile açıklayıcı faktör analizi sonuçlarının tutarlı olmasını sağlayacaktır. Öte yandan döndürme (rotation) yaklaşımı olarak ise promax döndürmenin tercih edilmesi doğrulayıcı faktör analizi ile tutarlı bulgular elde etmeye yardımcı olacaktır. Çünkü promax döndürme yönteminin de örneklem büyüklüğünün 150 üzerinde olduğu durumlarda verimli sonuçlar ürettiği bilinmektedir (Thompson, 2004; Costello ve Osborne, 2005). Nitekim doğrulayıcı faktör analizi için uygulanacak yapısal eşitlik modellemesi için de yüksek örneklem büyüklüğü kapsamında stabil ve sağlıklı kestirimler elde edilebildiği (Byrne, 2001) göz önünde bulundurulduğunda promax dön-

dürme yöntemi ile kovaryans temelli yapısal eşitlik modeli analizi sonuçlarının tutarlı olacağı öngörülebilir.

Toparlamak gerekirse; açıklayıcı faktör analizini uygularken çıkarılma (extraction) yöntemi olarak en büyük benzerlik yöntemi (maximum-likelihood method¹) tercih edilmediğinde doğrulayıcı faktör analizi sonuçları ile açıklayıcı faktör analizi sonuçları tutarlı olmayacaktır. Çünkü doğrulayıcı faktör analizinin uygulandığı kovaryans temelli yapısal eşitlik modeli analizi en büyük benzerlik yöntemi (maximum-likelihood method) ile kestirimlerde bulunmaktadır. Doğrulayıcı faktör analizi sonuçları ile açıklayıcı faktör analizi sonuçlarının olabildiğince tutarlı olabilmesi için en büyük benzerlik yöntemi (maximum-likelihood method) ve promax döndürme doğrultusunda açıklayıcı faktör analizinin gerçekleştirilmesi önerilmektedir (Costello ve Osborne, 2005).

(10) Yansıtıcı (reflective) ölçüm ve oluşturuvcu (formative) ölçüm nedir?

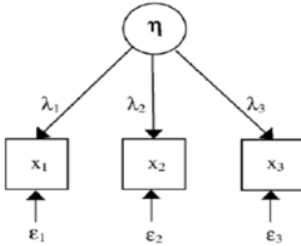
Davranışsal ve sosyal bilimler kapsamında bilimsel araştırma sürecinde psikometrik ölçüm araçları olan ölçeklerden sıklıkla yararlanılmaktadır. Psikometrik ölçekleri kullanmanın temel sebebi, araştırmacının ölçümlemeyi düşündüğü gizil yapıya (latent construct) ilişkin ölçüm yapabilmesidir. Bu psikometrik ölçüm gerçekleştirilirken araştırmacı tarafından ölçüm modelinin tasarlanması gerekir. Bir diğer deyişle, ölçümü yapılan gizil yapı ile ölçüm aracı olan ölçek ifadeleri arasındaki nedensellik ilişkisinin hangi yönde olduğuna araştırmacının ölçüm sürecinde karar vermesi gereklidir. Dolayısıyla psikometrik ölçekler kullanıma hazır unsurlar değil, ölçüm modelinin tasarlanarak uygulanması gereken ve teori ile ölçüm arasında köprü oluşturan birer metodolojik aletlerdir.

Özellikle pazarlama disiplinde yaygın ölçüm modelinin yansıtıcı (reflective) ölçüm modeli olduğu gözlemlenmektedir (Hair vd.,

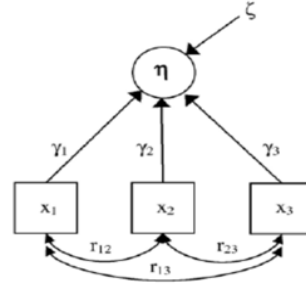
1 Maximum likelihood yönteminin uygulanabilmesi için normal dağılım varsayımını verinin ihlal etmemiş olması gerektiği şartı sağlanmalıdır.

2011). Hatta aslında oluşturucu (formative) ölçüm modeli kapsamında operasyonelleştirilmesi gereken birçok psikometrik ölçüm aracının hatalı şekilde yansıtıcı ölçüm modeli kapsamında operasyonelleştirildiği vurgulanmaktadır (Diamantopoulos ve Winklhofer, 2001). Bu noktada yansıtıcı (reflective) ölçüm ile oluşturucu (formative) ölçüm arasındaki farkın bilincinde olmak bir çözüm sunabilir. Yansıtıcı ölçüm, nedenselliğin gizil yapıdan ölçüm ifadelerine doğru olduğu ve gizil yapıdaki değişimin tüm ölçüm ifadeleri üzerinde değişime neden olacağı savunulan ölçüm modelidir. Oluşturucu ölçüm ise nedenselliğin ölçüm ifadelerinden gizil yapıya doğru olduğu ve gizil yapıdaki değişimin tüm ölçüm ifadeleri üzerinde değişime neden olmayacağı savunulan ölçüm modelidir (Bagozzi, 2011; MacKenzie vd., 2011). Aşağıda yer alan Şekil 2’de yansıtıcı ölçüm modeli ve oluşturucu ölçüm modeli görselleştirilmiştir.

Yansıtıcı ölçüm modeli (reflective)



Oluşturucu ölçüm modeli (formative)



Şekil 2: Yansıtıcı ve oluşturucu ölçüm modeli

Yukarıda yer alan Şekil 2’de görüldüğü üzere yansıtıcı ölçüm modeli kapsamında ölçüm ifadeleri hata terimine sahip iken oluşturucu ölçüm modeli kapsamında ise ölçüm ifadeleri hata terimine sahip değildir. Ayrıca yansıtıcı ölçüm modelinde ölçüm ifadelerinin ortak öncülü gizil yapı olduğundan dolayı yansıtıcı ölçüm modelinde ölçüm ifadeleri arasında ilişki (korelasyon) olması beklenir. Dolayısıyla yansıtıcı ölçüm modeline sahip bir ölçüm aracına ilişkin içsel tutarlılık katsayısı (Cronbach’s Alpha) hesaplanabilmektedir. Oluşturucu ölçüm modeli kapsamında ise ölçüm ifadeleri hata terimine sahip olmadığından dolayı ölçüm ifadeleri arasında ilişki (korelasyon) beklenmez. Dolayısıyla oluşturucu ölçüm modeline sahip bir ölçüm aracına ilişkin içsel tutarlılık katsayısı hesaplanamaz.

rucu ölçüm modelinde ise ölçüm ifadelerinin doğrusal toplamı gizil yapıyı oluşturmaktadır (Bagozzi, 1994). Bir diğer deyişle, oluşturucu ölçüm modelinde her bir ölçüm ifadesi gizil yapının farklı bir yönünü temsil etmektedir. Örneğin; sosyo-ekonomik statü (SES), gelir ve eğitimin bir kombinasyonu olarak kavramsallaştırıldığında SES gizil yapısı oluşturucu ölçüm modeli ile operasyonelleştirilmelidir. Çünkü sadece geliri çok yüksek olup eğitimi düşük bir kimse en üst SES grubuna ait olabilirken eğitimi ve geliri iyi düzeyde olan bir kimse de aynı SES grubuna ait olabilir. Hatta eğitimi çok yüksek olup geliri orta düzeyde olan bir kimse de bu SES grubuna ait olabilir. Görüldüğü üzere SES'in bileşenleri olan gelir ve eğitim ilişkili bir şekilde SES'i oluşturmamaktadır. Oluşturucu ölçüm modelini kavramanın bir yolu olarak da gizil yapının otomobil, ölçüm ifadelerinin ise bu otomobilin farklı parçaları olduğu senaryo üzerinden tartışmaktır. Arabanın bileşenleri olan kaporta, lastik, motor ve iç döşemeler bir araya gelerek otomobili oluştururken otomobilin kalitesinin bir birim daha iyileşmesi durumunda tüm bu bileşenlerde iyileşme beklemek yersizdir. Bu doğrultuda oluşturucu ölçüm modelini yansıtıcı ölçüm modelinden ayıran temel fark gizil yapı ile ölçüm ifadeleri arasındaki nedenselliğin yönüdür.

Yaygın olarak kovaryans temelli yapısal eşitlik modellemesi analizi uygulamaları sunan AMOS, LISREL ve MPLUS gibi paket programların varlığı davranışsal ve sosyal bilimlerdeki araştırmacıların ölçüm dendiğinde akıllarına sadece yansıtıcı ölçüm modelinin gelmesine sebep oluyor olabilir (Diamantopoulos ve Winklhofer, 2001). Özellikle ölçek geliştirme ve kullanım sürecinde ilk adım olarak içsel tutarlılık katsayısı olan Cronbach's Alpha katsayısının kontrol edilmesi geleneği de oluşturucu ölçüm modellerinin göz ardı edilmesine katkı sağlıyor olabilir. Bu noktada araştırmacılar ölçümlemeye çalıştıkları gizil yapının ilk olarak yansıtıcı ölçüm modeli ile mi yoksa oluşturucu ölçüm modeli ile mi operasyonelleştirileceğine karar veriyor olması gerekir. Nitekim oluşturucu ölçüm modeli kapsamında operasyonelleştirilen bir psikometrik ölçeğin Cronbach's Alpha katsayısının yeterince yüksek olmasını ($>.70$) beklemek yersiz olacaktır. Hatta oluşturucu ölçüm modeline sahip psikometrik ölçeklerin

Cronbach's Alpha katsayısının özellikle düşük olması beklenir. Çünkü doğrusal regresyon modeli ile test edilen oluşturucu ölçüm modeli analiz edilirken çoklu doğrusallık sorununun (multi-collinearity) olmadığı dikkatlice kontrol edilmelidir. Bir diğer deyişle, oluşturucu ölçüm modelindeki ölçüm ifadeleri mümkünse ilişkili olmamalıdır (Hair vd., 2016).

Ölçüm ifadeleri arasından bazı ifadelerin silinmesine ilişkin olarak da yansıtıcı ve oluşturucu ölçüm modelleri arasında farklılık bulunmaktadır. Yansıtıcı ölçüm modelinde ölçüm ifadeleri birbirlerinin yerlerini tutabilirken (interchangeable), oluşturucu ölçüm modelindeki ölçüm ifadeleri ise birbirlerinin yerlerini tutabilme özelliğine sahip değildir (Diamantopoulos vd., 2008). Çünkü yansıtıcı ölçüm ifadelerinin ilişkili olması beklenirken oluşturucu ölçüm ifadeleri ilişkisiz olması beklenmektedir. Bu yüzden oluşturucu ölçüm modelinde yer alan bir ölçüm ifadesinin elenmesi, ölçümlenmeye çalışılan gizil yapının bir kısmının artık ölçümlenememesi anlamına gelebilir. Dolayısıyla oluşturucu ölçüm modelleri kapsamında ölçüm ifadesi eleme sürecinde ekstra dikkatli olmak gereklidir.

Yansıtıcı ölçüm modelleri ile oluşturucu ölçüm modelleri analizleri açısından da farklılaşmaktadır. Yansıtıcı ölçüm modelleri, kovaryans temelli yapısal eşitlik modellemesi (AMOS, LISREL) ile analiz edilebilirken, oluşturucu ölçüm modelleri varyans temelli yapısal eşitlik modellemesi (PLS-SEM) ile analiz edilebilmektedir (Hair vd., 2011). Ancak oluşturucu ölçüm modellerini kovaryans temelli yapısal eşitlik modellemesi kapsamında MIMIC (Bagozzi vd., 1981) model yaklaşımı ile analiz edebilmek de mümkündür.

Toparlamak gerekirse, bir gizil yapı veya değişkenin ölçümünü yapmak üzere ölçek geliştirme veya kullanımı yapmadan önce ölçüm modeli araştırmacı tarafından kurgulanmalıdır. Bu noktada yansıtıcı ölçüm modeli ile oluşturucu ölçüm modeli arasındaki ayrımın bilincinde olmak kritik öneme sahiptir. Bu doğrultuda ölçüm ifadelerinin ilişkili olmasını yersiz bir şekilde beklemenin ve ölçüm modelinin yanlış şekilde belirlenmesinin (misspecification) önüne geçilmiş olacaktır.

SONUÇ

Bu yazıda, davranışsal ve sosyal bilimlerdeki araştırmacılar arasında yaygınca tartışıldığı düşünülen 10 metodolojik sorunun cevabı sunulmaya çalışılmıştır. Tartışma ortamında sunulan cevapların araştırmacıların zihninde yer alan sorulara cevap vermiş olması ümit edilmektedir. Ayrıca davranışsal ve sosyal bilim alanlarındaki ulusal araştırmacıların metodolojik sorularına cevap sunan ve devamında bu cevapların akademik dergilerde tartışıldığı yöntem köşelerinin (methodological corner) oluşturulması için de bu yazının bir girişim olarak değerlendirilmesi beklenmektedir. Bu yazı kapsamında kantitatif araştırma metodolojisi alanında psikometrik ölçüm merkezli sorulara cevap sunulmaya çalışılmıştır. Gelecekte kalitatif araştırma metodolojisine yönelik ve ekonometrik temele sahip kantitatif araştırma metodolojisine yönelik yaygınca sorgulanan soruların cevaplarını içeren yazıların yayınlanması ümit edilmektedir.

Davranışsal ve sosyal bilimlerdeki araştırmacıların metodolojik araçları kullanırken arka planda yatan felsefenin ne olduğunu kavraması ve ezberci uygulamadan uzak bir şekilde metodolojik uygulamaları gerçekleştirmesi ulusal literatürün metodolojik kalitesinin artmasına yardımcı olacaktır. Bu amaçla, araştırmacılar metodolojik sorunlarında nasıl-merkezli sorular sormak yerine neden-merkezli sorular sormalıdır. Çünkü nasıl-merkezli sorular istatistiksel uygulamaların hangi program ve hangi 'tıklamalar'dan faydalanılarak yapılmasını odak noktasına alırken, neden-merkezli sorular ise metodolojik ve istatistiksel uygulamaların ardında yatan gerekçelerin açığa çıkmasını sağlamaktadır. Böylece neden-merkezli sorular sorularak metodolojik ve istatistiksel konular daha iyi kavranabilir.

Ulusal alanda araştırmacılar tarafından uygulanan ve deneyimlenen araştırma süreçlerinin doğasını anlamak ve bu doğrultuda geliştirici öneriler sunmak teşvik edilmesi gereken bir girişimdir. Nitekim ulusal yönetim ve organizasyon alanı özelinde bilimsel araştırma yapısı şekline ilişkin yapılan tespitler (Özen, 2002) sadece araştırma metodolojisine ilişkin değil araştırma konusunun ele alınış biçimine

ilişkin de sorunlar olduđuna işaret etmektedir. Dolayısıyla teorinin ne olduđuna, teori geliştirme sürecine ve sağlam temellere sahip araştırma projeleri oluşturabilmeye yönelik konuların da gelecekte ulusal akademik dergilerde tartışılıyor olması önemlidir.

KAYNAKÇA

- Aguinis, H., Edwards, J. R. ve Bradley, K. J. (2016). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*, DOI: 10.1177/1094428115562749.
- Aiken, L. S. ve West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Los Angeles, CA: Sage.
- Altman, D. G. ve Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, 311, 485.
- Altunışık, R., Coşkun, R., Bayraktaroğlu, S. ve Yıldırım, E. (2010). *Sosyal bilimlerde araştırma yöntemleri (SPSS uygulamalı)*. Ankara, TR: Pegem Akademi.
- Armstrong, J. S. (1970). How to avoid exploratory research. *Journal of Advertising Research*, 10(August), 27-30.
- Aronson, E., Carlsmith, J. M. ve Ellsworth, P. C. (1990). *Methods of research in social psychology (2nd ed.)*. New York, NY: McGraw-Hill.
- Bagozzi, R. P. ve Fornell, C. (1982). Theoretical Concepts, Measurements, and Meaning. In Fornell C. (Ed.) *A Second Generation of Multivariate Analysis (Vol. 2)*, 24-38, New York, NY: Praeger Publishers.
- Bagozzi, R. P. ve Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science*, 40(1), 8-34.
- Bagozzi, R. P., Fornell, C. ve Larcker, D. F. (1981). Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*, 16(4), 437-454.
- Bagozzi, R. P. (1994). Measurement in Marketing Research: Basic Principles of Questionnaire Design in Principles of Marketing Research. In Bagozzi R. P. Bagozzi (Ed.), 1-49, Cambridge, MA: Basil Blackwell.
- Bagozzi, R. P. (2011). Measurement and meaning in information systems and organizational research: Methodological and philosophical foundations. *MIS Quarterly*, 35(2), 261-292.
- Barlett, J. E., Kotrlik, J. W. ve Higgins, C. C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1), 43-50.

- Baron, R. M. ve Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Baumeister, R. F., Vohs, K. D. ve Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16(6), 351-355.
- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, 66, 153-158.
- Bem, D. 2003. Writing the Empirical Journal Article. In Darley, J. M., Zanna, M. P. ve Roediger III, H. L. (Eds.), *The Compleat Academic: A Practical Guide for the Beginning Social Scientist (2nd ed.)*, Washington, DC: American Psychological Association.
- Berkowitz, L. ve Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37(3), 245-257.
- Burrell, G. ve Morgan, G. (1979). *Sociological paradigms and organisational analysis*. London, UK: Heinemann.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara, TR: Pegem Akademi Yayıncılık.
- Byrne, B. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York, NY: Springer Verlag.
- Byrne, B. M. (2001). Structural equation modeling: Perspectives on the present and the future. *International Journal of Testing*, 1(3-4), 327-334.
- Calder, B. J. ve Tybout, A. M. (1999). A vision of theory, research, and the future of business schools. *Journal of the Academy of Marketing Science*, 27(3), 359-366.
- Calder, B. J., Phillips, L. W. ve Tybout, A. M. (1981). Designing research for application. *Journal of Consumer Research*, 8(2), 197-207.
- Calder, B. J., Phillips, L. W. ve Tybout, A. M. (1982). The concept of external validity. *Journal of Consumer Research*, 9(3), 240-244.

- Campbell, D. T. ve Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Cohen, J., Cohen, P., West, S. G. ve Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.)*. Mahwah, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Costello, A. B. ve Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1-9.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.
- Çil, B. (2008). *İstatistik*, Ankara, TR: Detay Yayıncılık.
- Darlington, R. B. ve Hayes, A. F. (2017). *Regression analysis and linear models: Concepts, application, and implementation*. New York, NY: Guilford Press.
- DeVellis, R. (1991). *Scale development: Theory and applications (4th ed.)*. London, UK: Sage.
- Diamantopoulos, A. ve Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269-277.
- Diamantopoulos, A., Riefler, P. ve Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, 61(12), 1203-1218.
- Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Hampshire, UK: Palgrave Macmillan.
- Dobbins, G. H., Lane, I. M. ve Steiner, D. D. (1988). A note on the role of laboratory methodologies in applied behavioural research: Don't throw out the baby with the bath water. *Journal of Organizational Behavior*, 9(3), 281-286.
- Enis, B. M., Cox, K. K. ve Stafford, J. E. (1972). Students as subjects in consumer behavior experiments. *Journal of Marketing Research*, 9(1), 72-74.
- Espinosa, J. A. ve Ortinau, D. J. (2016). Debunking legendary beliefs about student samples in marketing research. *Journal of Business Research*, 69(8), 3149-3158.

- Faul, F., Erdfelder, E., Buchner, A. ve Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.
- Ferber, R. (1977). Research by Convenience. *Journal of Consumer Research*, 4(June), 57-58.
- Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society (B)*, 17, 69-77.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5), 587-606.
- Gonzalez, R. (2009). *Data analysis for experimental design*. New York, NY: Guilford Press.
- Gordon, M. E., Slade, L. A. ve Schmitt, N. (1986). The “science of the sophomore” revisited: From conjecture to empiricism. *Academy of Management Review*, 11(1), 191-207.
- Hair, J. F., Anderson, R. E., Tatham, R. L. ve Black, W. C. (1998). *Multivariate data analysis with readings (5th ed.)*, Englewood Cliffs, NJ: Prentice Hall.
- Hair, J. F., Ringle, C. M. ve Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, 19(2), 139-152.
- Hair, J. F., Hult, G. T. M., Ringle, C. M. ve Sarstedt, M. (2016). *A Primer on Partial Least Squares Structural Equation Modeling (2nd ed.)*. Thousand Oaks, CA: Sage.
- Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.
- Jarvis, C. B., MacKenzie, S. B. ve Podsakoff, P. M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, 30(2), 199-218.
- Job, V., Walton, G. M., Bernecker, K. ve Dweck, C. S. (2013). Beliefs about willpower determine the impact of glucose on self-control. *Proceedings of the National Academy of Sciences (PNAS)*, 110(37), 14837-14842.
- Kalaycı, Ş. (2005). *SPSS uygulamalı çok deđişkenli istatistik teknikleri*. Ankara, TR: Asil Yayın Dađıtım.

- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196-217.
- Kurtuluş, K. (2010). *Araştırma Yöntemleri*. İstanbul, TR: Türkmen Kitabevi.
- Lynch, J. G. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9(December), 225-239.
- Lynch, J. G. (1983). The role of external validity in theoretical research. *Journal of Consumer Research*, 10(1), 109-111.
- MacCallum, R. C., Browne, M. W. ve Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149.
- MacKenzie, S. B., Podsakoff, P. M. ve Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293-334.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... ve Laitin, D. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30-31.
- Munafo, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, ve Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- Neuman, W. L. (2002). *Social Research Methods (4th ed)*. Boston, MA: Allyn and Bacon.
- Neyman, J. (1950). *First course in probability and statistics*. New York, NY: Holt.
- Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological Methods* 5, 241-301.
- Nunnally, J. C. (1978). *Psychometric methods*. New York, NY: McGraw Hill.
- Nuzzo, R. (2014). Statistical errors. *Nature*, 506, 150-152.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

- Özen, Ş. (2002). Türkiye'deki örgütler/yönetim arařtırmalarında törenselleşme sorunu. *Yönetim Arařtırmaları Dergisi*, 2(2), 5-31.
- Patton, M. Q. (2005). *Qualitative Research. Encyclopedia of Statistics in Behavioral Science*. New Jersey, John Wiley & Sons.
- Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Routledge.
- Richins, M. L. ve Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, 19(3), 303-316.
- Rosenthal, R. ve Rosnow, R. L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York, NY: McGraw-Hill.
- Rosnow, R. L. ve Rosenthal, R. (1996). Computing contrasts, effect sizes, and counterfactuals on other people's published data: General procedures for research consumers. *Psychological Methods*, 1(4), 331-340.
- Simmons, J. P., Nelson, L. D. ve Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.
- Simonsohn, U., Nelson, L. D. ve Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534-547.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559-569.
- Steenkamp, J. B. E. ve Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-90.
- Sutton, R. ve Staw, B. M. 1995. What theory is not. *Administrative Science Quarterly*, 40, 371-384.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Trope, Y. ve Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440-463.

- Vandenberg, R. J. ve Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Wagenmakers E. J., Wetzels R., Borsboom D. ve van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, 100, 426-432.
- Wansink, B. ve Ittersum, K. (2016). Boundary research: Tools and rules to impact emerging fields. *Journal of Consumer Behaviour*, 15(5), 396-410.
- Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, 40(3), 385-390.
- Winer, R. S. (1999). Experimentation in the 21st Century: The Importance of External Validity. *Journal of the Academy of Marketing Science*, 27(3), 349-358.
- Zhao, X., Lynch, J. G. ve Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206.