



Çok Değişkenli İstatistiksel Analizler İçin Örneklem Büyüklüğü

Multivariate Statistical Analysis and Required Sample Size

Ömer AKBULUT¹
Cantürk ÇAPIK²

¹Giresun Üniversitesi, Fen Bilimleri Enstitüsü, Biyosüreç Mühendisliği Anabilim Dalı, Giresun, Türkiye
²Atatürk Üniversitesi, Hemşirelik Fakültesi, Halk Sağlığı Hemşireliği Anabilim Dalı, Erzurum, Türkiye

ÖZ

İstatistiksel tekniklerin çoğu örneklem büyüklüğüne duyarlı olduğu için veri analizinde örneklem büyüklüğü önemli bir yere sahiptir. Çok değişkenli analizlerde örneklem büyüklüğü için esnek bir yaklaşım söz konusu olmasına rağmen yine de araştırmacılar için örneklem büyüklüğünü doğru olarak belirlemek önemli bir problemdir. Çok değişkenli istatistik teknikler başta sağlık bilimleri olmak üzere, biyoloji, ekonomi, eğitim ve mühendislik olmak üzere birçok bilim alanında kullanılmaktadır. Bu derlemede özellikle sağlık bilimleri alanından örneklerle Çok Değişkenli Varyans Analizi (MANOVA), Çok Değişkenli Regresyon Analizi, Çok Değişkenli Lojistik Regresyon Analizi, Faktör Analizi, Temel Bileşenler Analizi, Yapısal Eşitlik Modeli, Kümeleme Analizi, Ayırma Analizi, Setler Arası Korelasyon Analizi, Çok Boyutlu Ölçekleme Analizi ve Bitişiklik Analizi ve teknikleri ele alınmıştır. Her bir analiz tekniği için gerekli örneklem büyüklüğü ölçütleri sunulmuştur.

Anahtar Kelimeler: Araştırma tasarımı, sağlık bilimleri, çok değişkenli istatistik analizler, örneklem büyüklüğü

ABSTRACT

Since most statistical techniques are sensitive to sample size, sample size has an important place in data analysis. There is a flexible approach for sample size in multivariate analysis; however, determining the sample size correctly is an important problem for researchers. Multivariate statistical techniques are used in many scientific fields, especially health sciences, biology, economics, education, and engineering. In this review, multivariate analysis of variance, multivariate regression analysis, multivariate logistic regression analysis, factor analysis, principal component analysis, structural equation modeling, cluster analysis, discriminant analysis, canonical correlation analysis, multidimensional scaling analysis, and conjoint analysis and techniques are discussed. The necessary sample size criteria for each analysis technique are presented.

Keywords: Experimental design, health science, multivariate statistical analysis, sample size

GİRİŞ

İstatistiksel analiz yöntemleri değişken boyutu ile tek değişkenli ve çok değişkenli olmak üzere iki grupta sınıflandırılırlar. Tek değişkenli istatistik analiz yöntemleri incelenen değişken dışındaki etkili/ilişkili tüm faktörleri katılımcılar/denekler için türdeş kabul eder.¹ Gerçekte ise incelenen değişken dışındaki faktörler bakımından deneklerin bir örnek veya türdeş olması çoğu zaman mümkün değildir. Bu durum ancak deneysel, kısmen de yarı deneysel çalışmalarda sağlanabilir. Ayrıca tek değişkenli istatistiksel analiz yöntemlerinin sahip olması gereken normallik, toplanabilirlik, varyansların homojenliği, kovaryansın sıfır olması gibi birçok varsayımları vardır.² Bu nedenlerle tek değişkenli analizler ile gerçek durumun aydınlatılması mümkün olmayabilir.

Çok değişkenli istatistiksel analizler, araştırılan olay ve ilgili olduğu değişkenleri dikkate alarak veya çok sayıdaki değişkeni daha az sayıda doğrusal faktörlerine indirgeyerek, değişkenler arasındaki karmaşık ilişkileri incelemek ve çözümlere ulaşmak için geliştirilmiş yöntemler bütünüdür.¹ Ayrıca tek değişkenli analiz yöntemleri için varsayımların gerçekleşmemesi durumunda çok değişkenli analizler gerçekçi bir yaklaşımdır. Bu nedenle çok değişkenli istatistiksel analizler veri çözümlemede önemli bir yere sahiptir.³

Çok değişkenli istatistiksel analizler daha karmaşık işlemler gerektirmekle birlikte, günümüzde bilgisayar yazılımları sayesinde yaygın ve etkili bir şekilde kullanılmaktadır. Çok değişkenli analiz yöntemleri

Geliş Tarihi/Received: 13.07.2021

Kabul Tarihi/Accepted: 08.04.2022

Sorumlu Yazar/Corresponding author:
Ömer AKBULUT
E-mail: omer.akbulut@giresun.edu.tr

Cite this article as: Akbulut Ö, Çapık C. Multivariate statistical analysis and required sample size. *J Nursology*. 2022;25(2):111-116.



Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

hemen hemen tüm bilim alanlarında ancak özellikle sağlık, psikoloji, biyoloji, sosyoloji, ekonomi, eğitim, mühendislik, tarım ve turizm alanlarında yaygın kullanılmaktadır.

Sağlık bilimlerinde klinik tanı ve prognozu belirlemek, ayrıca risk tahmini yapabilmek için çok değişkenli istatistiksel modellere ihtiyaç vardır. Bir bireyin teşhis edilmemiş bir hastalığa veya duruma sahip olma riskini tahmin etmek veya bir bireyin gelecekte belirli bir olayı yaşama riskini tahmin etmek için bu yöntemler kullanılabilir.⁴ Çok değişkenli analizler kullanılırken; doğru yöntem ve doğru değişken seçimi ile sonuçların doğru yorumlanması en önemli noktalar. Bu konudaki bir diğer önemli nokta ise örnekleme planı ve örneklem büyüklüğüdür.⁵

Yeterli örneklem büyüklüğüne karar verme genelde tüm araştırmacıların karşılaştıkları en önemli sorunlardan biridir. Bu sorunun çözümü düşünüldüğü kadar kolay olmayıp, bilgi, deneyim ve uzmanlık gerektirmektedir.⁶ Çünkü, örneklem büyüklüğünün saptanması çok etkenli bir işlemler bütünüdür.

Bu yazının ana konusu olan örneklem büyüklüğü, araştırmada kullanılacak araştırma tasarımı ve istatistiksel tekniklerinin doğası ile ilgilidir. Genellikle istatistiksel tekniklerin çoğu örneklem büyüklüğüne duyarlıdır. Çok değişkenli istatistiksel teknikler için gerekli örneklem tek değişkenli istatistik tekniklerin aksine daha pratik bir yaklaşımla belirlenmektedir. Yazının ileri ki bölümlerinde çok değişkenli analiz yöntemleri kısaca tanımlanarak, bu analizler için gerekli veya yeterli örneklem büyüklükleri açıklanacaktır. Örnekleme oluşturan nesnelere materyale ve çalışma alanına göre; birim, birey, ünite, denek, gözlem veya katılımcı olarak adlandırılmaktadır. Söz konusu bu kavramlar derlemede konu akışına göre eşanlamlı olarak kullanılacaktır.

Bu çalışmada özellikle sağlık bilimleri alanında çok değişkenli analizlerin kullanımına yönelik bazı açıklamaların ele alınmasının da faydalı olacağı öngörülmektedir. Daha geniş ve sosyal bir bakış açısıyla, örneğin ağırlık ölçekleri aynı zamanda, kemik/yağ/su oranı hesaplayan makinalar gibi, bazı tansiyon aletleri ve giyilebilir teknolojilerin bir kısmı sonuçları aslında mini bir "çok değişkenli" analiz yaparak verirler. Tatbiki bilgisayarda çalışan istatistik yazılımları kadar büyük akıllı bir teknoloji barındırmazlar ama mini bir yonga ile sadece kullandıkları alanda çok değişkenli çözümler sunmaktadırlar. Bu örnekten yola çıkarak bu derleme; çok değişkenli analizlerin bilgisayar ortamında yapılacağı var sayılarak özellikle örneklem büyüklüğüne yoğunlaşarak hazırlanmıştır.

Sağlık bilimleri literatüründe, çok değişkenli regresyon analizi^{7,8} çok değişkenli lojistik regresyon analizi⁹⁻¹¹ ve kümeleme analizinin¹² kullanıldığı çalışma örneklerine sıklıkla rastlanmaktadır. Ayrıca sağlık alanında çok değişkenli doğrulayıcı ve açıklayıcı faktör analizlerinin her geçen gün daha çok yaygınlaştığı bildirilmektedir.¹³ Çapık¹³ bir yapısal eşitlik modeli olan doğrulayıcı faktör analizinin hemşire akademisyenler tarafından bulguların doğrulanması ve bir ölçeğin faktör yapısını belirlemesi için kullanıldığını bildirmektedir. Araştırmacı¹³ bazı çalışmalarında doğrulayıcı faktör analizinin ölçek geliştirmede kullanıldığı örnekleri de sunmuştur.

Bu derlemenin özellikle araştırma yapan aktif bilim insanlarına ve doktora öğrencilerine katkı sağlayacağını, yüksek lisans öğrencilerine ise bir farkındalık oluşturabileceği düşünülmektedir.

Başlıca Çok Değişkenli İstatistik Analiz Yöntemleri

Çok değişkenli istatistik analiz yöntemleri bu konuda hazırlanmış temel kaynaklarda, Çok Değişkenli Varyans Analizi (MANOVA), Çok

Değişkenli Regresyon Analizi, Çok Değişkenli Lojistik Regresyon Analizi, Setler Arası Korelasyon Analizi, Temel Bileşenler Analizi, Faktör Analizi, Kümeleme Analizi, Ayırma Analizi, Çok Boyutlu Ölçekleme Analizi, Bitişiklik Analizi ve Yapısal Eşitlik Modeli olarak sayılmaktadır.¹⁴⁻¹⁷

Çok Değişkenli Varyans Analizi (Multivariate Analysis of Variance, MANOVA)

ANOVA tekniğinde bir bağımlı değişken üzerinde, bir veya birden fazla bağımsız değişkenin etkisi analiz edilmektedir. Bağımsız değişkenlerde grup sayısı iki olduğu zaman normal dağılımlı çok sayıdaki bağımlı değişkenlerin birlikte analizi Hotelling T² Testi ile yapılabilmektedir. Bağımsız değişken sayısı ikiden fazla veya bir değişkenin alt grup sayısı ikiden fazla ise Hotelling T² Testi kullanılmamaktadır.¹ Bu durumda yani bağımlı ve bağımsız değişkenlerin iki veya daha fazla olduğu durumlarda değişkenlerin birlikte analizi MANOVA ile yapılabilmektedir. Ancak burada bağımlı değişkenlerin dağılımının normal dağılıma uygun olması varsayımı güvenilir analiz sonuçları için gereklidir.

MANOVA analizinin anlaşılabilirliğini daha somut bir şekilde sağlamak için Rababah ve arkadaşlarının 2019 yılında yürüttükleri bir araştırma incelenebilir.¹⁸ Bu çalışmada üniversite öğrencilerinin sağlık okur-yazarlığı incelenmiştir. Sağlık okur-yazarlığı ölçeği (HLQ, Health Literacy Questionnaire) 9 alt boyutlu sıralı (ordinal) ve toplanabilir olmayan bir ölçektir. Öğrencilerin 8 ana ve 19 alt grupta sınıflandırılan sosyo-demografik özelliklerinin sağlık okur-yazarlığı ölçek puanlarına etkisini incelemiştir. Çalışmanın örnekleme 520 kişiden oluşmuştur. Araştırmada 9 bağımlı değişken ve 8 bağımsız değişken arasındaki örüntü SPSS yazılımında MANOVA modeli ile çözümlenmişler ve şu açıklamaları sunmuşlardır. Adı geçen yazılımının MANOVA testi için Pillai's Trace, Wilks' Lambda, Hotelling's Trace ve Roy's Largest Root olmak üzere dört farklı çıktı sunduğunu bildirmişlerdir. Araştırmacılar çözümlenmiş sonuçlarına Roy's Largest Root çıktıları ile karar vermişlerdir. Karar vermede bu tekniği seçmelerinin nedenleri olarak ise; 1. Bazı değişkenlerin ikiden fazla grup içermesi, 2. Gruplar arasındaki örneklem hacminin eşit olmaması ve 3. Dağılımın normal olması olarak açıklamışlardır.

MANOVA tekniğinin uygulanabilmesi için tüm bağımlı değişkenlerin örneklem boyutu aynı olmalıdır. Veri sayısı az olan değişkenler eksik gözlem gibi algılanacağından analiz dışı kalacaktır.

MANOVA için örneklem hacmi, tek değişkenli faktöriyel (çok faktörlü) analiz tekniğinde olduğu gibi bağımsız değişkenin en alt grubunda en az 15-20 arasında gözlem bulunmalıdır. Materyalin bir örnekliliği durumunda bu sayı bir miktar daha azaltılabilir. Örneğin A, B, C faktörlerini ve a, b, c faktörlerin seviye sayılarını göstermek üzere a=3, b=2, c=4 seviyeli bir faktöriyel tasarımda en alt grup sayısı (etken kombinasyonu sayısı) $3 \times 2 \times 4 = 24$ adettir. Her bir alt gruba 10 birey alınması durumunda örneklem hacmi $24 \times 10 = 240$ olacaktır. Deney tasarımları terminolojisinde faktörler alfabenin ilk ve büyük harfleri, seviye sayıları bu harflerin küçük yazılımları ile gösterilir.¹⁹

Çoklu Değişkenli Regresyon Analizi, ÇDRA (Multiple Regression Analysis, MRA)

Bir bağımlı Y ve bir bağımsız X değişkeni (tahmin edici) arasındaki fonksiyonel ilişki basit regresyon analizi ile incelenir. Bir bağımlı değişken Y ile birden fazla bağımsız değişken X_n (tahmin ediciler) arasındaki fonksiyonel ilişki ise çok değişkenli regresyon analizinin (ÇDRA) konusudur. ÇDRA sağlık bilimleri alanında yaygın kullanılan bir teknik olup, fonksiyonel ilişki; doğrusal veya doğrusal olmayan modellerle incelenebilmektedir.

Tablo 1. Çoklu Regresyon Analizinde Örneklem Büyüklüğü İçin Cohen'in Etki Büyüklükleri^{20,21}

İstatistiksel Test	İlgili Etki Büyüklüğü	Etki Büyüklüğü Değeri		
		Küçük	Orta	Büyük
Çok Değişkenli Regresyon Analizi	R^2	0,02	0,13	0,26
	f^2	0,02	0,15	0,35

ÇDRA için hem bağımlı hem de bağımsız değişkenlerin dağılımı normal dağılıma uygun olmalıdır. Eğer değişkenler sürekli ise öngörülen Tip I hata (α) testin gücü ($1 - \beta$) ve Cohen tarafından bu analiz için bildirilen aşağıdaki etki büyüklüğü^{20,21} değerlerine göre gerekli örneklem hesaplanır (Tablo 1). Bir diğer yaklaşım bağımsız değişkenlerin her biri için 15-20 gözlem sayısı olacak şekilde örneklem büyüklüğü belirlenir. Bu sayı minimum 10 deneğe kadar azaltılabilir. Örneğin 10 bağımsız değişken için gerekli örneklem hacmi bu teknik için $10 \times 20 = 200$ olacaktır. Ancak adım adım (stepwise) regresyon analizi için sonuçların genelleştirilebilirliği açısından daha fazla örneklem büyüklüğü gereklidir.¹⁷

ÇDRA uygulamalarında örneklem büyüklüğü için geliştirilmiş teknikler ve formüller de mevcuttur. Bu bağlamda Green²² iki farklı yaklaşım önermektedir. Eğer araştırmacı R^2 değerini önemsiyorsa, örneklem büyüklüğünü $n \geq 50 + 8(k)$ formülü ile hesaplayabilir. Burada k bağımsız değişken sayısıdır. Eğer araştırmacı β hatasını önemsiyorsa hesaplamada $n \geq 104 + (k)$ formülünü kullanabilir. Brooks and Barcikowsky²³ ÇDRA için farklı bağımlı değişken ve R^2 ölçülerine göre kendi geliştirdikleri formül ile diğer 6 yöntemi karşılaştırmalı olarak incelemişlerdir. Araştırmacıların sunduğu çizelgeden yararlanılarak 4 ve 15 bağımsız değişken için 7 yöntemle göre hesaplanmış örneklem büyüklüğü değerleri Tablo 2'de sunulmuştur. Yazarlar kendi metodlarını güç bakımından diğer metotlardan üstün olduğunu ifade etmişlerdir.²³ ÇDRA'inde örneklem büyüklüğü için bu Tablo 2 ve kaynaktan yararlanılabilir.

Çok Değişkenli Lojistik Regresyon Analizi, ÇDLRA (Multiple Logistic Regression Analysis, MLRA)

Tek veya çok değişkenli regresyon analizlerinde bağımlı değişken normal dağılımlı ve sürekli. Lojistik regresyon analizinde bağımlı değişken kategoriktir. Sınıflamalı veya sıralamalı ölçeklerden biri ile ölçülmüş veya sürekli değişken kategorik hale getirilmiş olabilir. Bağımsız değişkenler sürekli, kategorik (kesikli) veya her ikisinin karması olabilir. Lojistik regresyon analizinde bağımlı değişkenin iki seviyesi varsa Binary Lojistik Regresyon, bağımlı değişkenin ikiden fazla seviyesi varsa Multiple Lojistik Regresyon analizi olarak adlandırılır. Lojistik regresyonda bağımsız değişken sayısı ikiden fazladır.

Sağlık bilimlerinde özellikle sosyal bilimlerle daha ilişkili olan hemşirelik alanında regresyon analizinin kullanılması mümkündür. Soyut kavramları ölçmekte kullanılan ölçeklerin depresyonda/sağlam, hasta/sağlam gibi iki gruplu kesim noktası olması durumunda tüm bağımsız değişkenler dağılım türüne bakılmaksızın bu analizle incelenebilir.

Lojistik regresyonda her değişkenin alt grubu için 15-20 örneklem olmak üzere veya her değişken için 20 örneklem olacak şekilde örneklem büyüklüğü belirlenir. Örneğin 5 adet bağımsız değişken varsa $5 \times 20 = 100$ bireylik örneklem büyüklüğü gereklidir. Burada herhangi bir değişkenin kategori sayısı 7 ise örneklem büyüklüğü $7 \times 20 = 140$ olacaktır.¹⁷ Örnek olarak Ertürk ve Özmen tarafından 2018 yılında yapılan çalışmada²⁴ "Hemşirelerin Profesyonel Tutumlarını Yordayan Değişkenlerin Belirlenmesi" konu olarak alınmıştır. Söz konusu çalışmada 11 adet bağımsız değişken 271 birimlik bir örneklemle incelenmiştir. Yukarıdaki hesaplama göre $11 \times 20 = 220$ kişilik örneklem yeterli olan bu çalışma daha büyük bir örneklemle (271) yürütülmüştür ve daha güvenilir bulgulara ulaşılmıştır.

Faktör Analizi, FA (Factor Analysis FA)

Faktör analizi, değişken sayısını azaltmak veya yeni yapılar ortaya çıkartmak amacıyla yapılır. Bu analizle aynı yapıyı ölçen çok sayıda değişkenden, az sayıda ve tanımlanabilir nitelikte anlamlı değişkenler elde edilir.¹ Diğer bir anlatımla FA, aralarında korelasyon bulunan ve gözlemsel olarak elde edilen çok sayıda değişkeni aralarında korelasyon bulunmayan ve gözlemsel olmayan daha az sayıda değişkene dönüştürme işlemidir. Faktör analizi ile elde edilen bu yeni değişkenlere "faktör" denir. FA korelasyon ve kovaryans matrisinden yararlanılarak gerçekleştirilir. Eğer veri setindeki değişkenlerin varyansları birbirinden önemli ölçüde farklı ise faktör analizi kullanılır. Değişkenler arasındaki korelasyon matrisi birim matris veya değişkenler arasındaki korelasyonlar çok düşük ise faktör analizinin yapılması gereksizdir.

FA örneklem büyüklüğü önemli bir yere sahiptir ve araştırma öncesinde belirlenmesi gerekir. Öncelikle denek (birey) sayısı değişken sayısından fazla olmalıdır. FA'nde örneklem büyüklüğünü belirlemede farklı yaklaşımlar mevcuttur.¹⁷

Tablo 2. Çok Değişkenli Regresyon İçin Örneklem Büyüklükleri²³

Metot	Tahminci sayısı = 4			Tahminci sayısı = 15		
	$R^2 = 0,50$	$R^2 = 0,25$	$R^2 = 0,10$	$R^2 = 0,50$	$R^2 = 0,25$	$R^2 = 0,10$
Brooks & Barcikowsky	84	124	244	315	465	915
Park&Dudycha	66	93	173	214	292	524
Cohen (COHEN)	16	48	144	26	78	235
Gatsonis & Sampson (GS)	25	55	165	42	88	256
30 : 1 (NP30)	120	120	120	450	450	450
50 + 8k (COMBO)	82	82	82	170	170	170
15 : 1 (NP15)	60	60	60	225	225	225

(1) Brooks and Barcikowsky²³'den özetlenmiştir.

Genel bir kural olarak FA çalışmalarında örneklem büyüklüğü, değişken sayısının (soru sayısı, madde sayısı) 5 katı olarak önerilmektedir.^{25,26} Kline²⁷ ise, örneklem büyüklüğünü madde sayısının 10 katı (10 : 1) olmasını önermekte ve bu oranın en az 2 : 1 olması gerektiği belirtilmektedir.²⁸

Siddiqui¹⁷ Thompson'a²⁹ atfen FA'nde en az 200 bireylik bir örneklemi önermekle birlikte 400 ve daha fazla bireye sahip bir örneklem daha güvenli sonuçlar elde edilebileceğini bildirmektedir. Kline²⁷ ise faktör yapısının açık ve az sayıda olduğu durumlarda bu rakamın 100 olabileceğini, bununla birlikte aynı araştırmacı güvenilir faktörler çıkartmak için 200 bireylik örneklem olması gerektiğini, ayrıca daha iyi sonuçlar almak için daha büyük örneklem ile çalışmanın yararlı olacağını belirtilmektedir (Akt: 28). Orçan³⁰ 100 değişken için 300 bireylik örneklemi yeterli görmüş ancak kümeleme analizinde kullanılan modelin karmaşıklığı arttıkça örneklem de artacağına dikkat çekmiştir.

FA'nde veri yapısı bakımından örneklem büyüklüğünün uygunluğunu belirlemek için Kaiser-Meyer-Olkin (KMO) testi de kullanılmaktadır. Bu test, gözlenen korelasyon katsayılarının büyüklüğü ile kısmi korelasyon katsayılarının büyüklüğünü karşılaştırmakta ve KMO testi sonucunda, elde edilen değer yorumlanmaktadır. KMO değeri 0,50'den düşük olması halinde örneklem büyüklüğü ve veri yapısı yetersiz olarak yorumlanmakta ve analize devam edilemeyeceği belirtilmektedir. Örneklem büyüklüğü için 0,60 KMO değeri yeterli 0,90 ve üzeri KMO değerlerin mükemmel olarak yorumlanmaktadır.³¹

Özet olarak, FA'nde uygun örneklem büyüklüğü madde sayısına bağlıdır. Güvenilir faktörler elde edebilmek için örneklem büyüklüğünün 200'den ve madde sayısının 5 katından az olmaması gereklidir. Daha güvenilir sonuçlar için örneklem sayısı 400 bireye ve madde sayısının 10 katına kadar artırılabilir. FA için veri yapısı ve örneklem büyüklüğü KMO testi ile de belirlenebilmektedir.

Temel Bileşenler Analizi, TBA (Principal Component Analysis, PCA)

Temel bileşenler analizi TBA, aralarında korelasyon bulunan k adet değişkeni, aralarında korelasyon bulunmayan ve orijinal değişkenlerin doğrusal bileşenleri olan daha az sayıda (m adet) değişkenle ifade etme yöntemidir. TBA orijinal k adet değişkenin açıkladığı durumu daha az değişkenle (m<k) açıklamakta olup, yeni değişkenler orijinal değişkenlerin doğrusal bileşenidir. TBA;

- Veri indirgemesi
- Tahmin işlemi
- Verileri bazı istatistiksel yöntemlerle analiz edilebilecek forma sokma ve
- Veri sıkıştırma amaçlarıyla kullanılmaktadır.¹

TBA uygulamalarında literatürde örneklem büyüklüğü için öngörülen bir kurala rastlanılmamıştır. Ancak TBA daha çok ÇDRA, KA veya FA'nin bir ara adımını oluşturmaktadır.¹ Genel bir yaklaşımla TBA'nin örneklem büyüklüğü, ara adım olduğu analiz için ana işleme ait gerekli örneklem büyüklüğü düzeyinde olmalıdır.

Yapısal Eşitlik Modeli, YEM (Structural Equation Modeling, SEM)

Yapısal eşitlik modeli (YEM) çok sayıda bağımlı ve bağımsız değişkenler arasındaki ilişkileri modelleyerek, birçok analizi tek bir süreçte gerçekleştirerek sonuca ulaşmayı sağlayan çok değişkenli bir tekniktir. YEM bağımlılık ilişkilerini tahmin etmek için, varyans-kovaryans analizleri, ÇDRA ve FA gibi analizlerin birleşimi ile oluşan bir yöntemdir. Ayrıca değişkenler arasındaki nedensel

ilişkiler ve düz ilişkilerin (korelasyon) bir arada ele alındığı modellerin test edilmesi içinde kullanılmaktadır.^{5,25,27,29}

Siddiqui¹⁷ birçok kaynağa atfen yaptığı derleme çalışmasında YEM için 10-15 tahminleyici (değişken) için 200-400 arasında örneklem büyüklüğünün gerekli olduğunu bildirmektedir. Araştırmacı¹⁷ ayrıca şu durumları da not etmektedir. YEM için en az 100 tercihen 200 örneklem gerekli görülmektedir. Değişken sayısı 10'dan fazla ise 200'den az örneklem büyüklüğü, tahminleri kararsız hale getirmekte ve istatistiksel gücü zayıflatmaktadır. Ayrıca yazar, basit bir model kullanan minimum 50 değişkenli bir çalışma için 450 örneklem büyüklüğünün gerekli olduğunu not etmiştir. Çapık¹³ ise yapısal eşitlik modellemesinde madde sayısının yanı sıra, faktör sayısının da önemli bir etken olduğunu vurgulamış, faktörlere olan madde dağılımının kavramsal yapıya uygun olması için örneklem hacminin önemine işaret etmiştir. Örneklem hacmi büyüdükçe uyum indekslerinin ve faktörlere olan dağılımların daha iyi sonuçlar verdiğini belirtmiştir.

Ayırma Analizi, AA (Discriminant Analysis DA)

Ayırma analizi (AA) ayırt edici değişken gruplarını belirlemede ve hangi gruba ait olduğu bilinmeyen bir bireyin ait olduğu grubu belirlemede kullanılır. Genel anlamda A kitlesinin (popülasyon) k adet değişken içeren normal dağılımlı ve çok değişkenli X gözlem matrisinde mevcut grupları birbirinden ayıracak en iyi fonksiyonu bulmada kullanılan ve yeni bir gözlemi ait olduğu gruba atamayı sağlayan bir tekniktir.^{1,32,33}

Bireylerin gruplandırılmasında kullanılmaları nedeniyle ayırma analizi kümeleme analizi ile benzerlik göstermektedir. AA'nde analiz başlangıcında grup sayısı bilinmemekte ve analiz sürecinde değişmektedir. Araştırmacı bireylerin bu grupların hangisine ait olduğunu belirler. KA'nde ise analiz başlangıcından küme sayısı bilinmemektedir. Bu iki analiz arasındaki bir diğer farklılık ise ayırma analizinden elde edilen bilgiler gelecekte kullanılabilirliktedir. KA ile elde edilen bilgiler mevcut durumu göstermekte ve genel tanımlamalar yapmayı sağlamaktadır.

AA'nde örneklem büyüklüğü için genel yaklaşım her bir değişken sayısının 4 veya 5 katı birey olmasıdır.²⁵ Örneğin değişken sayısı 10 ise 50 bireylik bir örneklem yeterlidir. AA'nde gruplarda gözlem sayılarının farklı olması problem oluşturmaz. Ancak gruplarda gözlem sayısının eşit olması tercih edilir.³⁴

Kümeleme Analizi, KA (Cluster Analysis, CA)

Kümeleme analizi (KA) çok değişkenli bir istatistik yöntemidir. Kümeleme analizi bir veri setinin diğer bir ifade ile veri matrisinin birbiri ile benzer alt grup veya sınıflara ayırma işlemleri sürecidir. KA yapılan veri setlerinde verilerin doğal grupları veya gruplamaları kesin olarak bilinmez. KA'nin amacı doğal grupları bilinmeyen bir veri setindeki üniteleri k adet değişkenden hesaplanan benzerlik ölçülerini esas alarak homojen alt gruplara ayırmaktır. Kısaca KA benzer olanları farklı olanlardan ayırarak verileri gruplandırma amacıyla yapılan bir sınıflandırma yöntemidir.

Bu amaçlar çerçevesinde KA yaygın bir kullanım alanına sahiptir. Kümeleme analizi, sağlık bilimleri, fen bilimleri, mühendislik bilimleri ve sosyal bilimler olmak üzere tüm bilim dallarında kullanılmaktadır.

KA için gerekli örneklem büyüklüğü hakkında veya kümeleme için kullanılan değişkenler ile bireyler arasında kabul görmüş bir kural yoktur.¹⁷ Kümeleme analizinin kaç değişkenle ve kaç bireylik

örnekleme yapılmasını sınırlayan bir kural da yoktur. Bu durum kümeleme analizinin sınırlılıklarından biri olabilir. Ancak geçerli sonuçlar elde etmek için büyük örnekleme çalışılması önerilir. Bununla birlikte Çokluk ve ark.³⁵ kümeleme analizi için $n < 250$ olması durumunda hiyerarşik kümeleme yöntemlerini diğer durumda hiyerarşik olmayan kümeleme yöntemlerini önermektedir. Bu duruma göre hiyerarşik kümeleme analizi için örneklem hacmi $n \leq 250$, hiyerarşik kümeleme analizi için $n > 250$ alınabilir.⁵ Kümeleme analizinde küme sayısının belirlenmesi diğer önemli bir sorundur. Uzaklık katsayıları, ağaç grafiği veya hiyerarşik olmayan kümeleme yöntemlerinin birisi analiz sonuçlarına bakılarak seçilebilir.³⁶

Setler Arası Korelasyon Analizi, SAKA (Canonical Corelation Analysis, CCA)

Setler arası korelasyon analizi (SAKA) veya kanonik korelasyon n adet bağımsız X_n ve m adet bağımlı Y_n değişkenden oluşan iki değişken seti arasındaki ilişkinin ölçüsüdür. Setlerdeki değişken sayısı eşit ($n = m$) veya farklı ($n \neq m$) olabilir. Değişken setlerinin bağımlı-bağımsız olarak tanımlanması zorunluluğu da yoktur. Bu analizden değişkenler arasında çoklu bağlantı (multicolinearity) olmaması ve verilerin çok değişkenli normal dağılım göstermesi varsayımları altında doğru sonuç alınabilir. Korelasyon katsayısı -1 ile 1 arasında değer alırken setler arası korelasyon ölçüsü 0 ile 1 arasında değer almaktadır.

SAKA'nde her değişken için değişken sayısının 5 katı,¹³ ancak daha güvenilir sonuçlar elde etmek için değişken sayısının 10 katı¹⁷ örneklem büyüklüğü gereklidir. Örnek vermek gerekirse, ortodontik bir çalışmada 5 açığı ve 7 uzunluk ölçüsünden oluşan iki veri seti arasındaki söz konusu korelasyon için 120 deneğin $[(5+7) \times 10 = 120]$ ölçüsü alınmalıdır.

Çok Boyutlu Ölçekleme Analizi, ÇBÖA (Multidimensional Scaling Analysis, MSA)

ÇBÖA, n birey (gözlem) arasında gözlenen benzerlikler veya farklılıklardan oluşan uzaklık değerlerini kullanarak bu nesnelerin tek veya çok boyutlu uzaydaki konumlarını ve ilişkilerini göstermeyi amaçlayan çok değişkenli bir istatistiksel yöntemdir.^{37,38} Bu analiz herhangi bir dağılım varsayımına sahip değildir.³ Analizin sonucu uzaysal bir haritadır. ÇBÖA'nde birbirine yakın noktalar bireylerin benzerliğini uzak noktalar farklılığı göstermekte olup hem çok değişkenli hem de keşfedici bir tekniktir.³⁹

ÇBÖA'nin uygulanabilmesi için değişkenlerin ve özellikle bireylerin karşılaştırılabilir olması önemlidir. Bu analiz tekniğinde örneklem büyüklüğünün belirlenmesinde esnek bir yaklaşım söz konusudur. Bununla birlikte güvenilir sonuçların elde edilmesi için veri kümesindeki birey sayısının, boyut sayısının 4 katından daha fazla olması önerilir. Örneğin, iki boyutlu bir gösterim için en az 9 birimle çalışılması gereklidir.

Bitişiklik Analizi, BA (Conjoint Analysis)

Nitelikleri nicel olarak karşılaştıran daha çok ekonomi alanında uygulama alanı bulan tüketicilerin ürün veya hizmeti tercih nedenlerini ölçen çok değişkenli bir istatistiksel tekniktir. Bu analiz tekniği ile tüketici karakteristiklerinin belirlenmesi, işletmelerin hangi noktalar üzerinde yoğunlaşmaları gerektiği, pazarın genel eğilimleri, yeni ürün ve hizmetin pazara olası etkileri, fiyat oluşumu vb. konular incelenmekte,^{3,40,41} keza sağlık alanında da kullanılmaktadır.⁴⁰

BA veriler daha çok anket yöntemi ile belirlenir. Veriler için BA istatistiksel varsayımlardan daha çok kavramsal varsayımları gerekli

görür. Bu varsayımların birincisi, araştırılan mal veya hizmetin özellikleri açık tanımlı olması diğer ise ankete cevap verenlerin cevaplarının anlamlı olmasıdır.⁴⁰ Bu kavram İngilizce CONside ve JOINT kelimelerinin bileşiminden oluşmakta⁴⁰ ve Türkçe literatürde "ilişkilendirme analizi," "ilişkilerin analizi" veya "birliktelikler analizi" olarak ifade edilebilmektedir.^{30,31}

Bitişiklik analizi çalışmaları için örneklem büyüklüğü daha çok araştırmanın amacına bağlı olup genellikle 150-1200 arasında olması önerilir. Güçlü kantitatif araştırmalar için, alt grupların karşılaştırılması amaç değilse, anlamlı bir öngörü yapmak için 300 bireylik örneklem gereklidir. Eğer araştırmanın amacı alt grupların karşılaştırılması ise ve anlamlı farklılıkları belirlemek ise, her grup için en az 200 örneklem büyüklüğü gereklidir. Örneğin bireyler 4 gruptan oluşuyorsa veya 4 gruba ayrılacaksa örneklem büyüklüğü $4 \times 200 = 800$ bireyden oluşması gereklidir.¹⁷ Bununla birlikte Acar ve Sönmez⁴¹ Akaah ve Korgaonkar⁴² 'a atfen BA tekniği için örneklem büyüklüğünün 100'den az olabileceğini ve bu tekniğin sınırlı örneklem büyüklüğünde de temsil edici özelliğinin olduğunu bildirmektedir. Siddiqui¹⁷ bir işletmenin (market gibi) araştırma ve geliştirme hipotezleri için 30 ve 60 arasında bireyin cevaplarının yeterli olabileceğini bildirmektedir.

Türkçe literatürde Turanlı ve ark.⁴⁰ iki grupta 48, Acar ve Sönmez⁴¹ çok faktörlü ve indirgenmiş bir tasarımda 50 bireylik örneklem (katılımcı) ile çalışmalarını gerçekleştirmişlerdir.

Sonuç olarak BA uygulamasında 200-300 bireylik bir örneklem büyüklüğü yeterli olabilir. Ancak hedef kitlenin sınırlı olması durumunda bu sayı 50-100 arasında olması uygundur.

SONUÇ

Araştırma hipotezine doğru ve güvenilir cevap verebilecek veri analiz yönteminin belirlenmesi araştırmacıların karşılaştıkları en önemli problemlerden biridir. Bir diğer önemli problem ise analiz yöntemi için gerekli örneklem büyüklüğünün belirlenmesidir. Araştırma bulgularının doğru raporlanması bakımından bu iki husus önem arz etmektedir. Söz konusu her iki problemin çözülmesi istatistiksel bilgi, deneyim ve uzmanlık gerektirir. Veri analizinde kullanılan çok sayıda çok değişkenli analiz yöntemi mevcuttur ve her biri için farklı örneklem hacmi gereklidir. Bu derlemede yaygın kullanılan çok değişkenli analizler ele alınarak bu analizler için gerekli örneklem büyüklükleri irdelenmiştir.

MANOVA tekniğinde her bir alt grup için minimum 10 gözlem (birey) gereklidir. ÇDRA için her bir değişken için 10-15 gözleme ihtiyaç vardır. Bu sayı 10 bireye kadar indirgenebilir. ÇDRA için örneklem büyüklüğü k değişken sayısını göstermek üzere $n \geq 50 + 8(k)$ formülü ile de hesaplanabilir. ÇDLRA için her bir değişken için 15-20 gözlem gereklidir. SAKA'nde her değişken için değişken sayısının en az 5 katı, güvenilir sonuç için 10 katı örneklem büyüklüğü gereklidir. FA'nde uygun örneklem büyüklüğü madde sayısına bağlıdır. FA'nde örneklem büyüklüğünün 200'den ve madde sayısının 5 katından az olmaması gereklidir. Daha güvenilir sonuçlar için örneklem sayısı 400 bireye ve madde sayısının 10 katına kadar artırılabilir. TBA'nin örneklem büyüklüğü, ara adım olduğu analiz gerektirdiği örneklem büyüklüğü düzeyinde olmalıdır. YEM için en az 100 tercihen 200 örneklem gerekli görülmektedir. KA için gerekli örneklem büyüklüğü hakkında kabul görmüş bir kural yoktur. AA'nde örneklem büyüklüğü için genel yaklaşım her bir değişken sayısının 4 veya 5 katı birey olmasıdır. ÇBÖA tekniğinde örneklem büyüklüğünün belirlenmesinde esnek bir yaklaşım söz konusudur. Bununla veri kümesindeki birey sayısının,

boyut sayısının 4 katından daha fazla olması gerektiği bildirilmektedir. BA uygulamasında 200-300 bireylik bir örneklem büyüklüğü yeterli olabilir. Ancak hedef kitlenin sınırlı olması durumunda bu sayı 50-100 arasında olması uygundur.

Hakem Değerlendirmesi: Dış bağımsız.

Yazar Katkıları: Fikir – Ö.A.; Tasarım – Ö.A., C.Ç.; Denetleme – Ö.A., C.Ç.; Kaynaklar – Ö.A., C.Ç.; Veri Toplanması ve/veya İşlenmesi – Ö.A., C.Ç.; Analiz ve/veya Yorum – Ö.A., C.Ç.; Literatür Taraması – Ö.A., C.Ç.; Yazıyı Yazan – Ö.A., C.Ç.; Eleştirel İnceleme – Ö.A., C.Ç.

Çıkar Çatışması: Yazarlar, çıkar çatışması olmadığını beyan etmiştir.

Finansal Destek: Yazarlar, bu çalışma için finansal destek almadıklarını beyan etmişlerdir

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – Ö.A.; Design – Ö.A., C.Ç.; Supervision – Ö.A., C.Ç.; Materials – Ö.A., C.Ç.; Data Collection and/or Processing – Ö.A., C.Ç.; Analysis and/or Interpretation – Ö.A., C.Ç.; Literature Review – Ö.A., C.Ç.; Writing – Ö.A., C.Ç.; Critical Review – Ö.A., C.Ç.

Declaration of Interests: The authors declare that they have no competing interest.

Funding: The authors declare that this study had received no financial support.

KAYNAKLAR

- Özdamar K. Paket Programlar ile İstatistiksel Veri Analizi Çok Değişkenli Analizler 2. *Kaan Kitabevi*. 1999:137-496.
- Yıldız N, Bircan H, Akbulut Ö. *Araştırma ve Deneme Metotları (Problemler ve Çözümleri)*. Atatürk Üniversitesi; 2012:33-34.
- Baydemir M. Bilimsel araştırmalarda istatistiğin ve doğru yöntem seçimlerinin önemi. *Bitlis Eren Univ Sosyal Bilimler Derg*. 2021;10(1):29-34.
- Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part II - binary and time-to-event outcomes. *Stat Med*. 2019;38(7):1276-1296. [CrossRef]
- Çapık C. İstatistiksel güç analizi ve hemşirelik araştırmalarında kullanımı: Temel bilgiler. *Anadolu Hemşirelik Sağlık Bilimleri Derg*. 2014;17(4):268-274.
- Sümbüloğlu V, Sümbüloğlu K. *Klinik ve Saha Araştırmalarında Örneklem Yöntemleri ve Örneklem Büyüklüğü*. Alf Ofset Matbaacılık; 2005.
- Kumcağız H, Aydın Avcı İ, Caner Ş. Yaşam doyumu, postpartum depresyon ve özkiyim olasılığı arasındaki ilişki. *Anadolu Hemşirelik Sağlık Bilimleri Derg*. 2018;17(3):1-9.
- Dündar T, Özsoy S, Toptaş B, Aksu H. Hemşirelikte mesleki değerler ve etkileyen faktörler. *Ege Univ Hemşirelik Derg*. 2019;35(1):11-19.
- Özkan S, Yılmaz E. Öğrenci hemşirelerin genel sağlık düzeyi, stresle baş etme yöntemleri ve etkileyen faktörler. *Ege Univ Hemşirelik Yüksek Okulu Derg*. 2010;26(2):67-82.
- Taşhan ST, Sever D. Stria gravidarum ve ilişkili faktörler. *Anadolu Hemşirelik Sağlık Bilimleri Derg*. 2012;15(1):33-39.
- Çobanoğlu A, Alkanat HÖ. Hastanede yatan hastaların akılcı ilaç kullanımına yönelik bilgi ve davranışlarının incelenmesi. *Anadolu Hemşirelik Sağlık Bilimleri Derg*. 2019;22(1):33-40.
- Keskin G, Yıldırım GÖ. Hemşirelerin kişisel değerlerinin iş doyumlarının incelenmesi. *Ege Univ Hemşirelik Yüksek Okulu Derg*. 2006;22(1):119-132.
- Çapık C. Geçerlik ve güvenilirlik çalışmalarında doğrulayıcı faktör analizinin kullanımı. *Anadolu Hemşirelik Sağlık Bilimleri Derg*. 2014;17(3):196-205.
- Manly BFJ. *Multivariate Statistical Methods*. Chapman and Hall; 1994.
- Alpar R. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş. *Bağcıran Yayınevi*. 1997.
- Richarme M. *Eleven Multivariate Analysis Techniques*; 2002. Available at: <https://decisionanalyst.com/media/downloads/MultivariateAnalysisTechniques>. (Access Date: 02.03.2021).
- Siddiqui K. Heuristics for sample size determination in multivariate statistical techniques. *World Appl Sci J*. 2013;27(2):285-287.
- Rababah JA, Al-Hammouri MM, Drew BL, Aldalaykeh M. Health literacy: Exploring disparities among college students. *BMC Public Health*. 2019;19(1):1401. [CrossRef]
- Yıldız N, Bircan H. *Araştırma ve Deneme Metotları. Atatürk Üniversitesi*; 2010.
- Özçomak MS, Çebi K. İstatistiksel güç analizi, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi üzerine bir uygulama. *Atatürk Univ İktisadi İdari Bilimler Derg*. 2017;31(2):413-431.
- G*Power Manual. 2017. Available at: <https://www.psychologie.hhu.de/Mathematisch-NaturwissenschaftlicheFakultaet/Psychologie/AAP/gpower>. (Access Date: 01.02.2021).
- Green SB. How many subjects does it take do to a regression analysis. *Multivariate Behav Res*. 1991;26(3):499-510. [CrossRef]
- Brooks GP, Barcikowsky RS. A new sample size formula for regression. Annual Meeting of American Educational Research Association. New Orleans LA; 1994.
- Ertürk C, Özmen D. Hemşirelerin profesyonel tutumlarını yordayan değişkenlerin belirlenmesi. *Dokuz Eylül Univ Hemşirelik Fak Electron Derg*. 2018;11(3):191-199.
- Child D. *The Essentials of Factor Analysis*. Continuum; 2006.
- Doğan N, Başokçu T. İstatistik tutum ölçeği için uygulanan faktör analizi ve aşamalı kümeleme analizi sonuçlarının karşılaştırılması. *Eğitimde Psikol Ölçme Değerlendirme Derg*. 2010;1(2):65-71.
- Kline P. *An Easy Guide to Factor Analysis*. Routledge; 1994.
- Büyüköztürk Ş. Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram Uygulamada Eğitim Yönetimi Derg*. 2002;32:470-483.
- Thompson B. *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. American Psychological Association; 2004.
- Orçan F. Açıklayıcı ve doğrulayıcı faktör analizi: İlk hangisi kullanılmalı? *Eğitimde Psikol Ölçme Değerlendirme Derg*. 2018;9(4):413-421.
- Tavşancıl E. *Tutumların Ölçülmesi ve SPSS ile Veri Analizi*. Nobel Kitap; 2005:230.
- Atakan C, Karabulut İ. Derinliğe dayalı diskriminasyon. *S Ü Fen Edebiyat Fakültesi Fen Dergisi*. 2003;22:53-63.
- Çamdeviren H. *Lojistik Regresyon ve Diskriminant Analizi*. Doktora Tezi. Ankara Üniversitesi; 2000:89-91.
- Zavorka S, Perrett JJ. Minimum sample size considerations for two-group linear and quadratic discriminant analysis with rare populations. *Communications in Statistics - Simulation and Computation* 2014;43(7):1726-1739.
- Çokluk Ö, Şekercioğlu G, Büyüköztürk Ş. Sosyal Bilimler için Çok Değişkenli İstatistik SPSS ve LISREL Uygulamaları. 2. Baskı. Pegem Akademi Yayıncılık; 2012.
- Kalaycı Ş. *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Asil Yayın Dağıtım; 2010.
- Mead A. Review of the development of multidimensional scaling methods. *Statistician*. 1992;41(1):27-39. [CrossRef]
- Çelik S. Çok boyutlu ölçekleme analizi ile hayvancılık açısından Türkiye'de illerin sınıflandırılması. *Erciyes Univ Fen Bilimleri Enstitüsü Derg*. 2015;31(4):1-6.
- Gürçaylılar Yenidoğan T. Pazarlama araştırmalarında çok boyutlu ölçekleme analizi: Üniversite öğrencilerinin marka algısı üzerine bir araştırma. *Akdeniz İİBF Dergisi*. 2008;08(15):138-169.
- Turanlı M, Taşpınar Cengiz D, Işık M. Konjoint analizi ile gazete tercihlerini etkileyen faktörlerin belirlenmesi. *Istanbul Univ Ekonometri İstatistik Derg*. 2013;19:1-26.
- Acar E, Sönmez H. Konjoint analizi ve genç kadın tüketicilerin hazır giyim mağaza tercihlerini etkileyen unsurların incelenmesi. *Akad Sosyal Araştırmalar Derg*. 2015;3(12):278-295.
- Akaah IP, Korgaonkar PK. A conjoint investigation of the relative importance of risk relievers in direct marketing. *J Advertising Res*. 1988;28(4):38-44.