

AI-Assisted Data Analysis: Validity and Reliability Estimations with ChatGPT-4.0

Duygu KOÇAK

Alanya Alaaddin Keykubat University, Alanya-Turkey

Article History

Submitted: 10.01.2025

Accepted: 13.02.2025

Published Online: 30.04.2025

Keywords

Artificial Intelligence
Data Analysis
ChatGPT
Psychometric Properties

Abstract

Purpose: This study aims to examine the performance of an artificial intelligence application in data analysis based on evidence of construct validity and reliability. For this purpose, exploratory factor analysis, Cronbach's alpha, and item-total test correlation calculations were performed using both R and ChatGPT on datasets generated under varying sample sizes and item numbers. The results obtained from these two methods were then compared.

Design & Methodology: This is a simulative study in which data were artificially generated and various statistical techniques were tested and compared. The sample sizes were set at 250, 500, and 1000, with item numbers set to 10 and 20, and repeated 100 times for simulation purposes. The generated datasets were subjected to exploratory factor analysis, and estimations for KMO value, total explained variance, and the number of factors were obtained. For internal consistency, Cronbach's alpha was calculated, and for item discrimination, item-total test score correlation coefficients were computed using both R and ChatGPT. The results were compared descriptively.

Findings: The study found that the values obtained from ChatGPT and R were consistent with each other. Specifically, the KMO value, total explained variance, number of factors, and Cronbach's alpha were found to be similar. However, as the sample size increased and thus the data matrix expanded, discrepancies were observed between the estimates obtained by ChatGPT and R.

Implications & Suggestions: The estimations for KMO, total explained variance, number of factors, and Cronbach's alpha values obtained by ChatGPT were in agreement with those obtained using R. Artificial intelligence can be used in decision-making processes that rely on quantitative values, but it is advised not to rely entirely on AI for tasks requiring researcher judgment, such as naming factors. In this study, analyses using the AI application were performed at a single point in time; it should be noted that results may vary if conducted at different times. Additionally, this study only tested ChatGPT; future research could explore whether similar results are obtained with other AI applications.



DOI:10.29129/inujgse.1617323

Yapay Zekâ Destekli Veri Analizi: ChatGPT-4.0 ile Geçerlik ve Güvenirlik Kestirimleri

Duygu KOÇAK

Alaaddin Keykubat Üniversitesi, Alanya-Türkiye

Makale Geçmişi

Geliş: 10.01.2025
Kabul: 13.02.2025
Online Yayın: 30.04.2025

Anahtar Sözcükler

Yapay Zekâ
Veri Analizi
ChatGPT
Psikometrik Özellik

Öz

Amaç:Bu araştırmada yapay zekâ uygulamasının veri analizi yapmadaki performansı yapı geçerliği ve güvenirlik kanıtları üzerinden incelenmesi amaçlanmıştır. Bu amaç doğrultusunda farklı örneklem büyüklüğü ve madde sayısı koşullarında üretilen verilerde açımlayıcı faktör analizi Cronbach alfa ve madde toplam test korelasyonu hesaplamaları R programı ve ChatGPT ile yapılarak karşılaştırılmıştır.

Yöntem:Araştırma verilerin yapay olarak elde edilerek çeşitli istatistiksel tekniklerin test edilerek karşılaştırmaların yapıldığı simülatif bir araştırmadır. Araştırmada örneklem büyüklüğü 250, 500 ve 1000, madde sayısı 10 ve 20 olarak değişimlenerek 100 tekrar ile simülatif olarak üretilmiştir. Üretilen veri setlerine açımlayıcı faktör analizi yapılarak KMO değeri, açıklanan toplam varyans oranı ve faktör sayısı kestirimleri; iç tutarlılık anlamında güvenirlik için Cronbach alfa kestirimleri ve madde ayırt ediciliği için madde toplam test puanı korelasyon katsayısı hem R hem de ChatGPT ile ayrı ayrı yapılmıştır. Elde edilen bulgular betimsel olarak karşılaştırılmıştır.

Bulgular:Araştırmada ChatGPT ve R programından elde edilen değerlerin birbiri ile uyumlu olduğu bulgusuna erişilmiştir. Buna göre KMO değeri, açıklanan toplam varyans değeri, faktör sayısı ve Cronbach alfa değeri benzerdir. Örneklem büyüklüğü arttıkça dolayısıyla veri matrisi büyüdükçe ChatGPT'nin elde ettiği kestirimler ile R arasında fark olduğu görülmüştür.

Sonuçlar ve Öneriler:ChatGPT tarafından kestirilen KMO, açıklanan toplam varyans, faktör sayısı ve Cronbach alfa katsayısı değerleri R ile uyum göstermektedir. Niceliksel değerlere bağlı kararlarda yapay zekâ kullanılabilir ancak faktör isimlendirme gibi araştırmacı yargısı gerektiren durumlarda bütünüyle yapay zekaya güvenmemek gerekmektedir. Bu araştırmada yapay zekâ uygulamasında analizler tek zamanda yapılmıştır, farklı zamanlarda yapıldığında farklı sonuçlar verebileceği göz önünde bulundurulmalıdır. Yapay zekâ uygulamalarından yalnızca ChatGPT üzerinden denemeler yapılmıştır farklı uygulamalarda da benzer sonuçlar verip vermediği araştırılabilir.



GİRİŞ

Sosyal bilimler araştırmalarında, geçerlik ve güvenilirlik analizleri, ölçme ve değerlendirme süreçlerinin temel taşlarını oluşturmaktadır. Sosyal bilimler disiplinlerinde ele alınan değişkenler, genellikle doğrudan gözlemlenemeyen tutum, ilgi, algı gibi soyut ve duyuşsal niteliklere sahiptir. Bu tür özelliklerin ölçülmesi, büyük ölçüde ölçekler, anketler veya testler gibi dolaylı ölçüm araçlarıyla gerçekleştirilmektedir (Creswell, 2014). Ancak, bu araçlardan elde edilen verilerin anlamlı, güvenilir ve genellenebilir sonuçlar sunabilmesi için kullanılan ölçüm araçlarının geçerli ve güvenilir olması gerekmektedir (DeVellis, 2017). Bu durum, sosyal bilimlerde araştırmacıların geçerlik ve güvenilirlik arayışını sürekli olarak gündemde tutmasına neden olmaktadır.

Geçerlik, bir ölçüm aracının, ölçülmek istenen kavramı ne derece doğru ölçtüğünü ifade eder (Messick, 1995). Yapı geçerliği ise, bir ölçeğin teorik çerçeveye uygun olarak ölçmeyi hedeflediği yapıyı ne kadar başarılı bir şekilde ölçtüğünü değerlendiren en yaygın geçerlik türlerinden biridir (BandalosveFinney, 2018). Bu bağlamda, yapı geçerliğini belirlemek için en sık kullanılan yöntemlerden biri açımlayıcı faktör analizidir (TabachnickveFidell, 2019). Bu analiz yöntemi, ölçme aracının alt boyutlarını belirlemeye, maddelerin hangi faktörlerle ilişkilendirilebileceğini anlamaya ve ölçeğin teorik temellerini doğrulamaya olanak tanır. Güvenirlik ise, bir ölçüm aracının tutarlılığını ve tekrarlanabilirliğini ifade eder (Nunnallyve Bernstein, 1994). Sosyal bilimlerde, güvenilirlik analizinde en yaygın kullanılan yöntemlerden biri olan Cronbach alfa katsayısı, bir ölçeğin iç tutarlılığını değerlendirmek için kritik bir göstergedir (TavakolveDennick, 2011). Cronbach alfa, ölçeğin tüm maddelerinin aynı yapıyı ölçme derecesini anlamaya yardımcı olur ve psikometrik analizlerde temel bir referans noktasıdır.

Arbağ ve Ertekin (2020), de Türkiye’de sosyal bilimler alanında gerçekleştirilen 751 lisansüstü tezi incelemiş ve bu tezlerde kullanılan güvenilirlik kestirim yöntemlerinden en yaygın olanının Cronbach alfa katsayısı, geçerlik türlerinden en sık tercih edileninin ise yapı geçerliği olduğunu tespit etmişlerdir. Bu durum, sosyal bilimlerde kullanılan ölçüm araçlarının psikometrik açıdan uygunluğunu değerlendirmek için geçerlik ve güvenilirlik analizlerinin ne derece merkezi bir rol oynadığını göstermektedir. Sosyal bilimler alanında kullanılan ölçeklerin geçerlik ve güvenilirliklerini değerlendirmek, bu disiplinlerde gerçekleştirilen araştırmaların temel gerekliliklerinden biridir. Bir ölçme aracının geçerli ve güvenilir olup olmadığını belirlemek için kullanılan analizler, hem metodolojik hem de uygulamalı açıdan kritik bir rol oynamaktadır. Sosyoloji, eğitim, psikoloji, psikiyatri ve ekonomi gibi birçok alanda, ölçme araçlarının uygunluğunu değerlendirmek için özellikle açımlayıcı faktör analizi ve Cronbach alfa katsayısı gibi yöntemler yaygın olarak tercih edilmektedir. Bu analizler, ölçme araçlarının hem geçerlik hem de güvenilirlik açısından güçlü bir temele sahip olmasını sağlarken, elde edilen verilerin doğruluğunu ve genellenebilirliğini artırmaktadır.

Faktör Analizi

Sosyoloji, eğitim, psikoloji, psikiyatri ve ekonomi gibi birçok alanda ölçekler, araştırmalarda sıkça kullanılmakta ve bu ölçeklerin geçerlikleri faktör analizi ile değerlendirilmektedir. Faktör analizi, bir ölçeğin ölçüm yapısının değerlendirilmesi için kullanılan güçlü bir yöntemdir ve verilerin uygunluğunun değerlendirilmesi, faktör çıkarımı, faktör döndürme ve yorumlama olmak üzere üç ana adımdan oluşmaktadır (Shrestha, 2021).

Veri setinin faktör analizine uygunluğunu belirlemek için bazı temel kriterlerin sağlanması gereklidir. Örneklem büyüklüğü ve maddeler arasındaki ilişkinin gücü, bu kriterlerin başında gelmektedir (TabachnickveFidell, 2019). Maddeler arasındaki ilişkinin gücünü değerlendirmek için korelasyon matrisinde 0,30’un üzerinde korelasyon katsayılarının bulunması beklenir (Arafat, Chowdhury, Qusarve Hafez, 2016). Bunun yanı sıra, çoklu bağıntının varlığı da önemli bir husustur; korelasyon katsayısı $r > 0,80$ olan değişkenler analizden çıkarılabilir.

Verilerin faktörleşmeye uygunluğunu değerlendirmek için genellikle Kaiser-Meyer-Olkin (KMO) Örneklem Yeterliliği Testi ve Bartlett'in Küresellik Testi kullanılmaktadır (Beavers, Lounsbury, Richards, Huck, ve Skolits, 2013). KMO değerlerinin ≥ 0.70 olması, verilerin faktör analizi için yeterli olduğunu gösterirken, Bartlett'in küresellik testinin anlamlı olması, analiz edilen verilerin birim matris olmadığını ve faktör analizi için uygun olduğunu belirtmektedir (Howard, 2016).

Faktör analizi, çok sayıda ilişkili değişkenden, daha az sayıda kuramsal olarak anlamlı değişkene, yani faktörlere indirgeme yapmayı amaçlar (Pallant, 2010). Veri setinin kaç faktör ile açıklanacağını belirlemek için Kaiser Kriteri (K1 yöntemi) gibi yöntemler kullanılabilir; bu yöntem, özdeğeri 1'in üzerinde olan faktörlerin belirlenmesine dayanır (Shrestha, 2021; Braekenve Van Assen, 2017). Bunun yanı sıra, belirlenen faktörler tarafından açıklanan toplam varyans oranı da faktör analizinde önemli bir göstergedir. Toplam varyans oranı 0,40 ile 0,60 arasında olduğunda, faktör analizi açısından uygunluk sağlamaktadır (Çokluk, Şekercioğlu ve Büyüköztürk, 2014).

Maddelerin hangi faktörlere yükleneceği, faktör döndürme ve madde faktör yük değerlerinin analizi ile belirlenir. Hair, Black, Babin ve Anderson (2014), madde faktör yüklerinin 0,30-0,40 arasında olması durumunda yorumlama için minimum düzeyi karşıladığını, 0,50 ya da daha büyük bir değer olduğunda ise uygulamada anlamlı kabul edildiğini belirtmiştir. 0,70'i aşan faktör yükleri ise iyi tanımlanmış bir yapının göstergesi olarak değerlendirilir. Ancak, binişik maddeler söz konusu olduğunda karar vermek güçleşebilir; bu durumda hem kuramsal yapı göz önünde bulundurulmalı hem de döndürme teknikleri uygulanmalıdır.

Cronbach Alfa Katsayısı

Bir ölçeğin güvenilirliğini değerlendirmek için kullanılan yöntemler arasında Cronbach alfa katsayısı, hem eğitim hem de psikoloji alanında en sık tercih edilen yöntemlerden biridir (Kartal ve Dirlik, 2016). İlk kez Lee J. Cronbach tarafından tanıtılan bu katsayı, bir ölçek veya test içindeki maddelerin iç tutarlılığını ölçer ve araştırmalarda standart bir yöntem haline gelmiştir (Cronbach, 1951). Cronbach alfa, bir ölçeğin maddelerinin birbirleriyle uyumlu olup olmadığını ve aynı yapıyı ne kadar iyi ölçtüğünü değerlendirir. Katsayı 0 ile 1 arasında değişir; genellikle 0.7'nin üzerindeki değerler kabul edilebilir olarak değerlendirilirken, yüksek risk taşıyan testlerde bu sınır 0.9'un üzerinde olmalıdır (Nunnallyve Bernstein, 1994; George ve Mallery, 2003). Ancak, 0.95'in üzerindeki çok yüksek değerler, ölçek maddelerinin gereksiz şekilde birbirini tekrarladığına ve benzersiz bilgi sağlamadığına işaret edebilir (Streiner, 2003).

Araştırmacılar, düşük korelasyon gösteren maddeleri çıkararak veya ifadeleri revize ederek, Cronbach alfa değerini artırabilirler. Sonuç olarak, Cronbach alfa katsayısı, ölçeklerin iç tutarlılık anlamında güvenilirliğini değerlendirmede en yaygın kullanılan araçlardan biri olarak kabul edilmektedir (DeVellis, 2017).

Madde Toplam Test Puanı Korelasyonu

Bir ölçek içindeki bireysel maddelerin, genel yapıya ne kadar katkı sağladığını değerlendirmek için kullanılan bir diğer ölçüt ise madde-toplam korelasyonudur. Bu korelasyon, belirli bir maddenin, testin ölçmeyi hedeflediği genel yapı ile ne kadar tutarlı olduğunu yansıtır (Field, 2009).

Madde-toplam korelasyonunun yüksek olması, maddenin genel ölçekle uyum içinde olduğunu ve hedeflenen yapının ölçümüne olumlu katkıda bulunduğunu gösterir (DeVellis, 2017). Genellikle 0.3'ün üzerindeki madde-toplam korelasyonu değerleri kabul edilebilir olarak değerlendirilir (Nunnallyve Bernstein, 1994). Düşük madde-toplam korelasyonları, maddenin ölçeğin genel yapısını ölçmedeki yetersizliğine işaret edebilir ve bu durum, Cronbach alfa katsayısında da düşüşe neden olabilir (Pallant, 2005).

Araştırmacılar, madde-toplam korelasyonunu ölçek geliştirme sürecinde, uyumsuz maddeleri belirlemek ve bu maddeleri gözden geçirmek veya çıkarmak için kullanmaktadır (Streiner, 2003). Madde-toplam korelasyonu, ölçeklerin rafine edilmesi ve hedeflenen yapının daha tutarlı bir şekilde ölçülmesini sağlamak için önemli bir araçtır (DeVellis, 2017).

Sosyal bilimler araştırmalarında kullanılan geçerlik ve güvenirlik analizleri, araştırma süreçlerinin vazgeçilmez bir parçası olmakla birlikte, bu analizlerin doğru bir şekilde gerçekleştirilmesi genellikle uzmanlık gerektiren bir süreçtir. Araştırmacıların, bu tür analizleri gerçekleştirebilmek için hem istatistiksel yöntemler konusunda bilgi sahibi olmaları hem de SPSS, R, Python gibi istatistik programlarını kullanabilme becerisine sahip olmaları gerekmektedir (Tabachnick ve Fidell, 2019; Field, 2009). Bununla birlikte, bu bilgi ve beceriler, özellikle sosyal bilimler alanında uzmanlaşmış ancak istatistiksel yöntemler konusunda sınırlı deneyime sahip araştırmacılar için zorlayıcı olabilmektedir (Beavers, Lounsbury, Richards, Huck, ve Skolits, 2013). Araştırmacıların bu süreçte yaptıkları hatalar, istatistiksel analizlerin sonuçlarının yanlış yorumlanmasına ve bilimsel literatüre katkının sınırlı kalmasına neden olabilir (Sharma, 1996).

İstatistiksel analizler, yalnızca analizlerin teknik kısmını öğrenmekle sınırlı kalmamakta; aynı zamanda bu süreçlerin mantığını kavramayı ve doğru yorum yapmayı da gerektirmektedir. Ancak, birçok araştırmacı, veri analizlerinin karmaşıklığı ve istatistiksel yazılım kullanımı ile ilgili teknik zorluklar nedeniyle bu süreçte ek yardım arayışına girmektedir (DeVellis, 2017; George ve Mallery, 2003). Özellikle, istatistik programlarının öğrenilmesi ve kullanılması zaman alıcı bir süreç olabilmekte, bu da araştırmacılar için hem maddi hem de manevi bir yük yaratmaktadır (TavakolveDennick, 2011).

Bu bağlamda, yapay zekâ tabanlı sohbet robotları, araştırmacıların istatistiksel analiz süreçlerindeki bilgi ve beceri eksikliğini gidermek için güçlü bir çözüm sunma potansiyeline sahiptir. Yapay zeka destekli araçlar, özellikle karmaşık istatistiksel analizlerin hızlı ve doğru bir şekilde gerçekleştirilmesi için giderek daha fazla tercih edilmektedir (Rudolph ve Tan, 2023; Kasneci, Kasneci, Stedevina, 2023). ChatGPT gibi modeller, hem istatistiksel yöntemlerin nasıl uygulanacağı konusunda rehberlik sağlamakta hem de kullanıcıların programlama bilgisi olmadan analiz yapmalarına olanak tanımaktadır (Mhlanga, 2023).

Özellikle, yapay zekâ sohbet robotlarının doğal dil işleme yetenekleri sayesinde, araştırmacılar yalnızca doğal dil kullanarak karmaşık istatistiksel analiz taleplerini ifade edebilir ve hızlı bir şekilde sonuçlara ulaşabilirler (Vaswani, 2017). Bu araçlar, yalnızca analiz süreçlerini hızlandırmakla kalmamakta, aynı zamanda kullanıcı hatalarını azaltarak analiz sonuçlarının doğruluğunu artırmaktadır (Baidoo-Anu ve Owusu-Ansah, 2023). Rudolph ve Tan (2023), yapay zekâ tabanlı araçların dil işleme yeteneklerinin, istatistiksel analiz süreçlerini kolaylaştırarak araştırmacıların iş yükünü hafiflettiğini ve daha verimli bir bilimsel süreç sunduğunu vurgulamaktadır.

Yapay zekâ sohbet robotları, yalnızca bir istatistiksel analiz aracı olarak değil, aynı zamanda bir öğrenme ve rehberlik platformu olarak da işlev görmektedir. Bu araçlar, karmaşık veri analizlerini daha erişilebilir hale getirerek, hem deneyimli hem de istatistiksel bilgiye daha az hakim olan araştırmacılar için önemli bir destek mekanizması sunmaktadır (Singh, Chen, Singhania, Nanavati ve Gupta, 2022). Bunun sonucunda, sosyal bilimlerdeki araştırma süreçleri, yapay zekâ destekli araçların kullanımıyla daha verimli, doğru ve erişilebilir bir yapıya dönüşmektedir.

Yapay zekâ (YZ) teknolojilerinin hızlı gelişimi, yalnızca teknoloji dünyasında değil, eğitim, sağlık, sosyal bilimler ve iş dünyası gibi birçok alanda önemli dönüşümlere yol açmıştır. Bu süreçte yapay zekâ destekli araçların kullanımı giderek yaygınlaşmakta ve araştırmacılar için çeşitli kolaylıklar sunmaktadır. Özellikle, dil işleme yetenekleri ile öne çıkan ChatGPT gibi modeller, son yıllarda büyük bir ilgi odağı haline gelmiştir (Rudolph & Tan, 2023; Baidoo-Anu ve Owusu-Ansah, 2023).

Yapay zekânın tarihsel gelişimi, 1950'lerde Alan Turing'in "Makineler düşünebilir mi?" sorusuyla başlayan bir süreçle şekillenmiştir. Daha sonra derin öğrenme, doğal dil işleme (NLP) ve makine öğrenmesi gibi alanlardaki ilerlemelerle, yapay zekâ teknolojilerinde çığır açan gelişmeler yaşanmıştır

(Vaswani, 2017). Özellikle GenerativePre-trainedTransformer (GPT) modellerinin geliştirilmesi, yapay zekâ uygulamalarının dil işleme yeteneklerini büyük ölçüde artırmış ve karmaşık analiz süreçlerini dahi gerçekleştirebilecek bir yapıya dönüşmüştür (Liang, Xiao, Liu, Zhou, Wang veLi, 2023).

GPT tabanlı modeller, büyük metin veri kümeleri üzerinde eğitilerek, dil üretiminde insan benzeri performans sergileyebilmektedir (Kasneci, Kasneci, StedevNaumann, 2023). Son yıllarda hızla gelişen bu yapay zekâ yeteneklerine veri analizi yapabilme kapasitesi de eklenmiştir. Bu durum, özellikle sosyal bilimlerde sıkça kullanılan istatistiksel analiz yöntemlerinin erişilebilirliğini artırma ve kolaylaştırma potansiyelini ortaya koymaktadır.

ChatGPT, Transformer mimarisi üzerine inşa edilmiş olup, veri işleme ve analiz etme yeteneklerine sahiptir (Vaswani, 2017). Transformer mimarisinin temelinde yer alan dikkat mekanizması (attentionmechanism), modelin girdi verisindeki önemli bölümlere odaklanarak analizleri daha doğru ve etkili bir şekilde gerçekleştirmesini sağlamaktadır (Tlili, Huang, Gao ve Kinshuk, 2023). Örneğin, bir veri analizi yapılırken, dikkat mekanizması sayesinde model, kritik verilere odaklanarak analiz süreçlerini optimize edebilir.

ChatGPT ayrıca otoregresif (autoregressive) bir model yapısına sahiptir. Bu yapı, modelin çıktığı tahmin ederken önceki girdilere dayanmasını sağlar ve talimatların sıralı bir şekilde işlenmesine olanak tanır (Dehouche, 2021). Örneğin, bir hipotez testi talebi alındığında, ChatGPT veriyi işler, uygun istatistiksel testi seçer ve sonuçları kullanıcıya sunar. Bu yetenek, ChatGPT'yi yalnızca metin üretimi için değil, aynı zamanda veri analizleri için de güçlü bir araç haline getirmektedir.

ChatGPT'nin veri analizlerindeki başarısının arkasında, denetimli öğrenme ve pekiştirmeli öğrenme teknikleri yatmaktadır. Özellikle İnsan Geri Bildirimiyle Pekiştirmeli Öğrenme (Reinforcement Learning from Human Feedback - RLHF), modelin kullanıcı geri bildirimlerine dayanarak performansını artırmasını sağlar (Singh, Chen, Singhania, NanavativeGupta, 2022). Bu öğrenme süreci, ChatGPT'nin yalnızca doğru analizler yapmasını değil, aynı zamanda analiz sonuçlarını etkili bir şekilde yorumlamasını mümkün kılmaktadır (Kasneci,Kasneci, StedevNaumann, 2023).

ChatGPT, eğitim sürecinde geniş bir metin veri kümesi üzerinde eğitilmiştir. Bu süreçte istatistik, matematik, veri bilimi ve programlama gibi alanlardan gelen verilerle beslenmiştir (Tlili, Huang, Gao ve Kinshuk, 2023). Böylece, tanımlayıcı istatistikler, hipotez testleri, regresyon analizleri ve faktör analizi gibi karmaşık istatistiksel teknikleri kolaylıkla kullanabilmektedir (Mhlanga, 2023). Ayrıca, ChatGPT'nin doğal dil işleme yetenekleri sayesinde, kullanıcılar programlama dillerine ihtiyaç duymadan yalnızca doğal dil kullanarak analiz taleplerini ifade edebilir ve sonuçlara ulaşabilirler (Rudolph & Tan, 2023).

ChatGPT'nin istatistiksel analiz yapabilme kapasitesi, güçlü algoritmalarına dayanmaktadır. Transformer mimarisi, dikkat mekanizmaları, otoregresif yapısı ve pekiştirmeli öğrenme teknikleri, ChatGPT'yi istatistiksel analizlerde etkili bir araç haline getirmektedir (Baidoo-AnuveOwusuAnsah, 2023). Bu teknolojik altyapı, araştırmacıların istatistiksel analiz süreçlerini hızlandırarak, bilimsel çalışmalarda yapay zekâdan daha etkin bir şekilde faydalanmalarını mümkün kılmaktadır.

Özellikle GPT-4'ün kullanıma sunulmasıyla, ChatGPT'nin daha karmaşık analizleri daha doğru bir şekilde gerçekleştirme kapasitesi artmıştır ve bu durum, araştırmacılar için büyük bir kolaylık sağlamaktadır (Terwiesch, 2023). Son yıllarda, Auto-GPT ve LangChain gibi projeler ortaya çıkmıştır. Auto-GPT, ChatGPT'nin GPT-4 tabanlı sürümüyle kendi başına görevler gerçekleştirebilen bir yapay zekâ aracı olarak dikkat çekmektedir. Python gibi programlama dillerini kullanarak veriyi işleyebilir ve analiz yapabilir. Ancak, bu tür araçların bile genellikle kullanıcı girdilerine ve doğru yapılandırmaya ihtiyaç duyduğu unutulmamalıdır.

Tüm bu gelişmelere rağmen, yapay zekâ kullanımının getirdiği etik kaygılar ve doğruluk sorunsalları hâlâ tartışılmaktadır (Nemurin, GillaniveWood, 2023; Gillani, Eynon, ChiabautveFinkel, 2023). Yapay zekânın doğru, güvenilir ve bilimsel bilgilere dayalı analizler sunup sunmadığı konusundaki belirsizlikler, araştırmacılar ve kullanıcılar için önemli bir sorgulama alanı yaratmaktadır. Bu nedenle, yapay zekâ

kullanılarak gerçekleştirilen veri analizlerinin doğruluğunu ve güvenilirliğini değerlendirmek, araştırmacılar açısından kritik bir öneme sahiptir.

Araştırmanın Amacı ve Önem

Sosyal bilimler alanında yapılan araştırmalar, genellikle doğrudan gözlemlenemeyen ve soyut niteliklere sahip değişkenleri ele almaktadır. Tutum, ilgi, algı gibi soyut kavramların ölçülmesi, bu alanın temel gereksinimlerinden biridir ve yalnızca doğru ölçme araçlarının kullanılmasıyla mümkün hale gelir (Creswell, 2014; DeVellis, 2017). Bu tür kavramların doğru bir şekilde ölçülmesi, hem güvenilir hem de geçerli sonuçlara ulaşmak için kritik bir öneme sahiptir. Bu nedenle, kullanılan ölçme araçlarının geçerliliğinin ve güvenilirliğinin belirlenmesi, sosyal bilimler araştırmalarında metodolojik bir zorunluluk olarak karşımıza çıkmaktadır.

Ölçme araçlarının geçerlik ve güvenilirlik düzeylerini belirlemek için sıklıkla başvurulan yöntemler arasında, yapı geçerliğini değerlendirmek amacıyla açımlayıcı faktör analizi, iç tutarlılığı ölçmek için Cronbach alfa katsayısı ve maddelerin ayırt edicilik düzeylerini belirlemek için madde-toplam test puanı korelasyon katsayısı yer almaktadır (Tabachnick ve Fidell, 2019; Tavakoli ve Dennick, 2011). Bu yöntemler, sosyal bilimler literatüründe, araştırma sonuçlarının doğruluğunu ve genellenebilirliğini artırmak amacıyla en sık kullanılan istatistiksel yöntemler arasında yer almaktadır. Ancak, bu analizlerin doğru bir şekilde uygulanabilmesi genellikle ileri düzeyde istatistiksel bilgiye ve bu analizleri gerçekleştirebilecek yazılım becerilerine sahip olmayı gerektirir (Field, 2009).

Son yıllarda, istatistiksel analizlerin yapay zekâ tabanlı uygulamalar aracılığıyla da gerçekleştirilebilmesi, bu alanda önemli bir yeniliği beraberinde getirmiştir. Yapay zekâ teknolojileri, özellikle karmaşık veri analizlerini otomatikleştirme ve hızlandırma potansiyeline sahiptir (Vaswani, 2017; Rudolph & Tan, 2023). Yapay zekâ destekli araçlar, istatistiksel bilgi ve yazılım becerileri gereksinimini azaltarak, analiz süreçlerini daha erişilebilir hale getirme potansiyeli taşımaktadır. Ancak, yapay zekâ araçlarının akademik ve profesyonel araştırma süreçlerinde ne derece güvenilir ve doğru sonuçlar sunduğu sorusu, hâlâ literatürde yeterince araştırılmamıştır (Kasneci, Kasneci, StederveNaumann, 2023; Singh, Chen, Singhanian, Nanavati ve Gupta, 2022).

Yapay zekâ tabanlı sohbet robotlarının, insan eliyle yapılan analizlerle ne ölçüde örtüştüğü veya farklılaştığı sorusu, özellikle veri analiz süreçlerinde önemli bir merak konusudur. ChatGPT gibi modellerin, açımlayıcı faktör analizi, Cronbach alfa hesaplaması ve madde ayırt edicilik indeksi gibi sosyal bilimlerde yaygın olarak kullanılan yöntemlerdeki performansı, bu bağlamda değerlendirilmeye ihtiyaç duyan bir konudur (Mhlana, 2023; Baidoo-Anu ve Owusu-Ansah, 2023).

Bu araştırmanın amacı, ChatGPT-4.0'ın bir veri analiz aracı olarak performansını değerlendirmektir. Araştırmada, yapay olarak üretilmiş verilere açımlayıcı faktör analizi, Cronbach alfa katsayısı ve madde ayırt edicilik indeksi hesaplamaları uygulanmış ve elde edilen sonuçlar R programı ile karşılaştırılmıştır. Bu kıyaslama, ChatGPT'nin geleneksel istatistik paket programlarına kıyasla doğruluk düzeyini ve güvenilirliğini değerlendirme fırsatı sunmuştur. Bu kapsamda, araştırma, ChatGPT'nin sosyal bilimlerde yaygın olarak kullanılan istatistiksel analizlerdeki performansını ortaya koymayı ve bu tür yapay zekâ araçlarının gelecekteki potansiyel kullanım alanlarını daha iyi anlamayı amaçlamaktadır.

YÖNTEM

Desen

Bu araştırmada yapay olarak üretilen veriler üzerinde yapay zekâChatGPT uygulamasının faktör analizi performansı incelenmiştir. Araştırmada simülatif veri kullanılarak bir program ve bir yazılımın performansları karşılaştırmalı olarak incelendiği için araştırma bu yönüyle simülasyon çalışması ile uyumludur (Dooley, 2002).

Verilerin Üretilmesi

ChatGPT'nin veri analizindeki performansının faktör analizi, güvenilirlik analizi ve madde ayırtedicilik indeksi üzerinden incelendiği bu çalışmada veriler R programı aracılığıyla üretilmiştir. Faktör analizi sürecinde örneklem büyüklüğünün sonuçlar üzerinde etkili olduğu ve 200'ün altındaki örneklem büyüklüğünün yetersiz, 300'ün iyi, 500'ün çok iyi ve 1000'in mükemmel olduğu yönündeki görüşler göz önünde bulundurulmuştur (Comreyve Lee, 1992; Gorsuch, 1983; TabachnickveFidell, 2007). Üretilen verilerde örneklem büyüklüğü 1000, 500 ve 250 olarak manipüle edilmiştir. Üretilen verilerde madde sayısı ise eğitim ve psikoloji alanında kullanılan ölçeklerdeki madde sayıları göz önüne önünde bulundurularak 10 ve 20 olacak şekilde sabitlenmiştir. Buna göre 20 ve 10 maddelik, 1000, 500ve 250 örneklem büyüklüğünde yanıt kategorisi 7-6-5-4-3-2-1 olacak şekilde 10 farklı veri seti yapay olarak 100 tekrar ile R programında üretilmiştir.

Verilerin Analizi

Veri üretiminin ardından üretilen verilere açımlayıcı faktör analizi, iç tutarlılık anlamında güvenilirlik belirleme yöntemlerinden Cronbach alfa ve maddelerin ayırtediciliği R yazılımı ve ChatGPT 4.0 ile ayrı ayrı olacak şekilde gerçekleştirilmiştir. R Development Core Team (2011) "psych" (RevelleveRevelle, 2015) paketi kullanılarak Kaiser-Meyer-Olkin katsayısı, Bartlett küresellik testi, Kaiser 1 kriteri ile faktör sayısı belirlenmiş, açıklanan toplam varyans oranları ve madde faktör yükleri hesaplanmış,maddelerin faktörlere dağılımı belirlenmiş, maddelerin ayırtedicilik indeksleri ve ölçeğin bütünü için Cronbach alfa katsayısı hesaplanmıştır.

Bu analizler hem R programında hem deChatGPT 4o ile gerçekleştirilmiştir. Açımlayıcı faktör analizi kesitimlerinde maksimum olabilirlik yöntemi temel alınmıştır.R ve Chat GPT ile elde edilen analiz bulguları karşılaştırılarak betimsel olarak incelenmiştir.

BULGULAR

Bu Araştırmada ChatGPT'nin veri analizi performansı açımlayıcı faktör analizi bulguları üzerinden R programı referans olarak incelenmiştir. Aşağıda karşılaştırmaya ilişkin açımlayıcı faktör analizi bulguları sunulmuştur.

Açımlayıcı faktör analizi parametrelerine ilişkin bulgular

Tablo 1.

Madde sayısı 20 olduğunda açımlayıcı faktör analizi bulguları

n		KMO	Bartlett	Faktör sayısı	Açıklanan varyans
1000	R	0,915	10585,87	5	67,84
	ChatGPT	0,914	10585,87	5	67,84
500	R	0.887	4749.93	5	65,81
	ChatGPT	0.887	4749.93	5	65,82
250	R	0.859	2204,47	6	69.55
	ChatGPT	0.859	2204,47	6	69.56

Tablo 1 incelendiğinde tüm örneklem büyüklüklerinde elde edilen KMO değerlerinin, faktör sayısının ve açıklanan toplam varyans oranının eşit olduğu görülmektedir. Buna göre madde sayısının 20 olduğu koşulda küçük ve büyük örneklem büyüklüklerinde ChatGPT'nin açımlayıcı faktör analizi performansının R ile uyumlu olduğu bulgusuna erilmiştir.

Tablo 2.

Madde sayısı 10 olduğunda açımlayıcı faktör analizi bulguları

n		KMO	Bartlett	Faktör sayısı	Açıklanan varyans
1000	R	0.841	5380.24	3	73.83
	ChatGPT	0.841	214.75	3	73.84
500	R	0.823	2423.03	3	71.85
	ChatGPT	0.823	131.13	3	71.85
250	R	0.809	1167.05	3	70.98
	ChatGPT	0.809	54.47	3	70.99

Tablo 2 incelendiğinde madde sayısı 20 olduğunda farklı örneklem büyüklüklerinde Chat GPT'nin sunduğu açımlayıcı faktör analizi bulgularının R ile uyumlu olduğu görülmektedir. Açıklanan varyans oranlarında örneklem büyüklüğü 1500, 1000 ve 250 olduğunda küsuratlarında 0.01'lik bir fark olduğu görülmektedir ancak bunun program ve uygulamanın yuvarlama kuralından kaynaklandığı düşünülmektedir ve büyük bir fark olmadığı ifade edilebilir. Tablo 1 ve Tablo 2 incelendiğinde örneklem büyüklüğü 1000, 500 ve 250, madde sayısı 10 ve 20 olduğu simülasyon koşullarında Chat GPT'nin faktör analizi performansının KMO, Bartlett küresellik testi, faktör sayısı ve açıklanan toplam varyans oranı açısından R paketi ile uyumlu sonuçlar verdiği görülmektedir.

Açımlayıcı faktör analizinde önemli aşamalardan biri de maddelerin faktörlere dağılımı ve madde faktör yük dağılımlarıdır. ChatGPT ve R programı ile yapılan açımlayıcı faktör analizleri sonucunda Tablo 1 ve Tablo 2'de aynı sayıda faktör belirlendiği görülmektedir. Maddelerin bu faktörlere dağılımları ise Tablo 3'te sunulmuştur:

Tablo 3.
Maddelerin faktörlere dağılımları (K=10)

		Maddeler									
n		1	2	3	4	5	6	7	8	9	10
1000	R	1	1	1	1	2	2	2	2	3	3
	ChatGPT	1	1	1	1	2	2	2	2	3	3
500	R	1	1	1	1	2	2	2	2	3	3
	ChatGPT	1	1	1	1	2	2	2	2	3	3
250	R	1	1	1	1	2	2	2	1	3	3
	ChatGPT	1	1	1	1	2	2	2	2	3	3

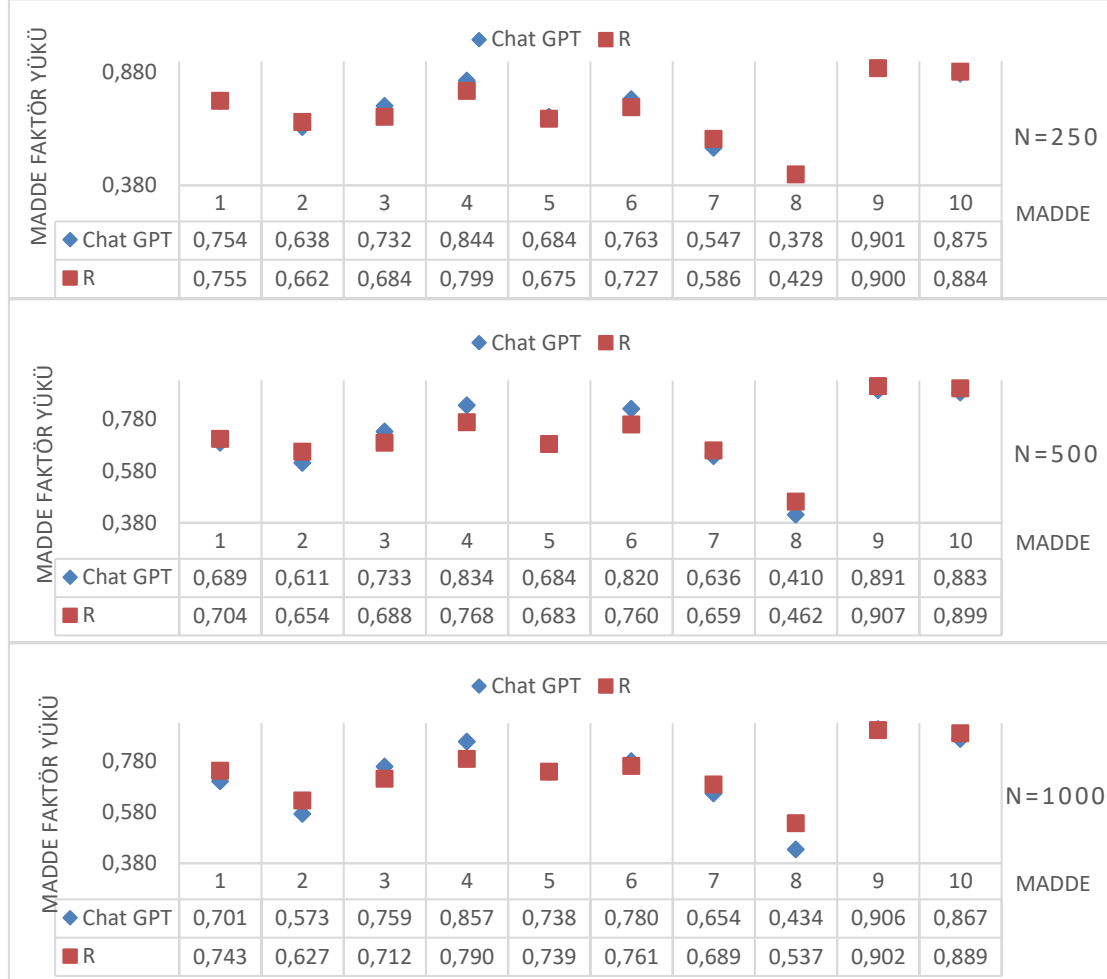
Tablo 3 incelendiğinde maddelerin faktörlere dağılımının örneklem büyüklüğü 1000, 500 ve 250 olduğunda uyumlu olduğu görülmektedir. Örneklem büyüklüğü 250 olduğunda ise bir maddenin farklı boyutlarda konumlandığı görülmektedir. Buna göre her ne kadar boyut sayısı tutarlı olsa da madde faktör dağılımlarının daima tutarlı olmadığı göz önünde bulundurulmalıdır.

Tablo 4.
Madde sayısı 20 olduğunda madde faktör dağılımları

N	250		500		1000	
Madde	R	ChatGPT	R	ChatGPT	R	ChatGPT
1	1	1	2	2	2	2
2	1	1	2	2	2	2
3	1	1	2	2	2	2
4	1	1	2	2	2	2
5	3	3	4	4	4	4
6	3	3	4	4	4	4
7	3	3	4	4	4	4
8	3	1	4	4	4	4
9	2	2	3	3	3	3
10	2	2	3	3	3	3
11	2	2	1	3	1	3
12	2	2	1	1	1	1
13	4	4	1	1	1	1
14	4	4	1	1	1	1
15	4	4	1	1	1	1
16	5	5	5	5	3	5
17	5	5	5	5	3	5
18	5	5	5	5	3	5
19	6	6	3	4	5	4
20	6	6	3	4	5	4

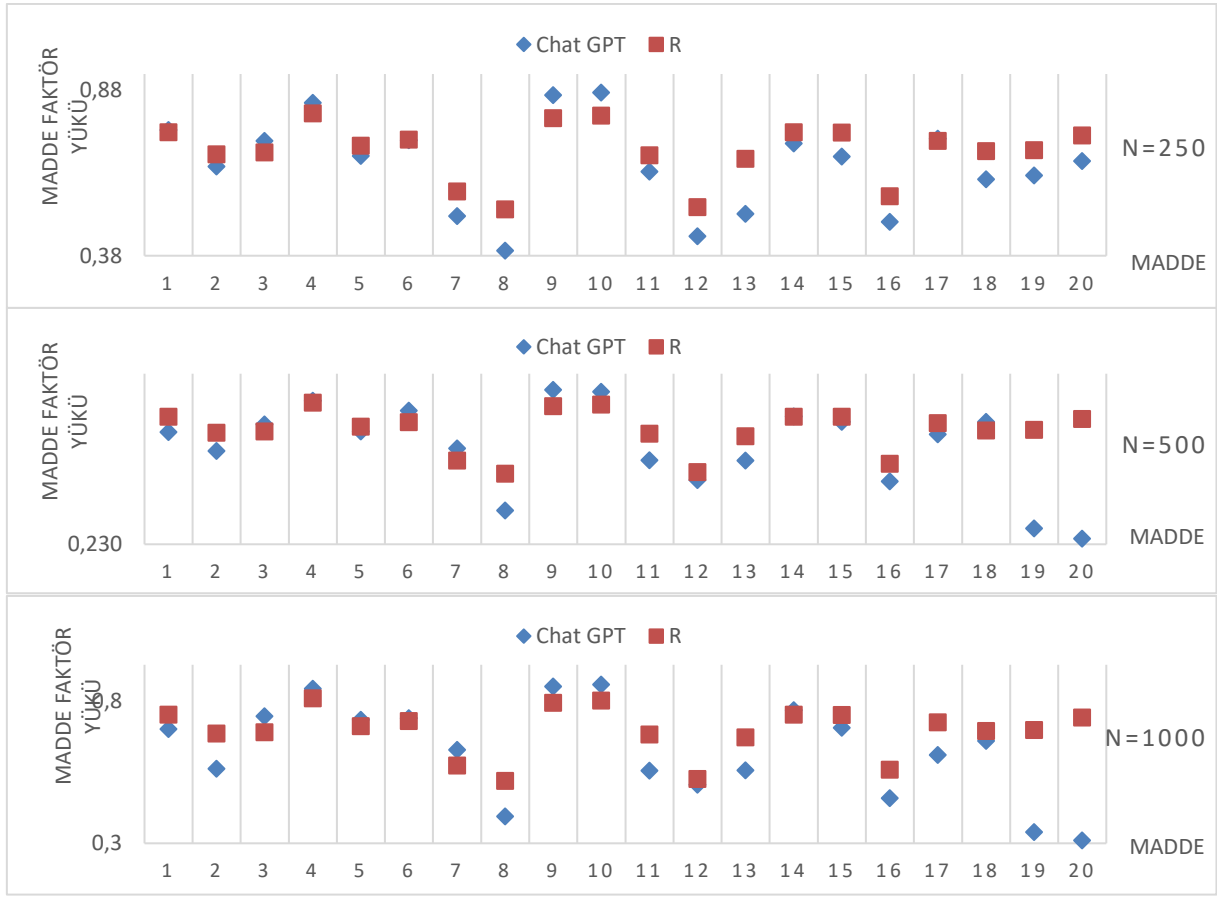
Tablo 4 incelendiğinde madde sayısı 20 iken özellikle 500 ve 1000 örneklem büyüklüklerinde maddelerin faktörlere dağılımları arasında farklılıklar olduğu görülmektedir. Buna göre madde sayısı artınca her örneklem büyüklüğünde R ve ChatGPT arasında madde faktör dağılımı açısından uyumsuzluklar olduğu ifade edilebilir. Tablo 3 ve Tablo 4 birlikte değerlendirildiğinde madde sayısı 10 olarak sabitlendiği koşulda yalnızca 250 örneklem büyüklüğünde bir madde için R ve Chat GPT arasında uyumsuzluk gözlenmiştir. Madde sayısı 20 olduğunda ise tüm örneklem büyüklüklerinde madde faktör

dağılımları arasında uyumsuzlukların olduğu, örneklem büyüklüğü arttıkça dolayısıyla veri matrisi büyükçe uyumsuzlukların arttığı gözlenmiştir.



Şekil 1. Madde faktör yükleri (K=10)

Şekil 1’de R ve Chat GPT ile hesaplanarak elde edilen faktör yükleri her bir madde için yer almaktadır. Bazı maddelerin faktör yükleri örtüşürken bazı maddelerin ise farklı kestirildiği görülmektedir. Örneğin örneklem büyüklüğü 1000 iken madde 8’in yükü R ile 0.537, ChatGPT ile ise 0.434 kestirilmiştir aralarında yaklaşık 0.1 fark bulunmaktadır. Benzer durum diğer örneklem büyüklüklerinde de gözlenmektedir.



Şekil 2. Madde faktör yükleri (K=20)

Şekil 2 incelendiğinde R ve ChatGPT ile kestirilen madde faktör yükleri arasında pek çok madde için büyük farkların olduğu görülmektedir. Her ne kadar her örneklem büyüklüğünde bazı maddelere ilişkin R ve ChatGPT kestirimleri arasında önemli farklar olmasa da maddelerin çoğunluğunda farkın büyük olduğu söylenebilir. R ya da ChatGPT için daha küçük veya daha büyük kestirimler sunuyor demek mümkün değildir çünkü bazı maddelerde R, bazı maddelerde ChatGPT daha yüksek kestirimler sunmaktadır.

Madde ayırteciiliği hesaplamalarına ilişkin bulgular

Üretilen verilerde simüle edilen koşullar birer ölçek verisi olarak ele alınıp maddelerin madde toplam test puanı korelasyonu hesaplaması ile ayırteciilikleri hesaplanmıştır.

Tablo 5.
Madde ayırt edicilik indeksleri (K=10)

n		1	2	3	4	5	6	7	8	9	10
1000	R	0.601	0.601	0.659	0.675	0.603	0.622	0.607	0.617	0.524	0.543
	ChatGPT	0.600	0.601	0.658	0.674	0.603	0.633	0.606	0.616	0.523	0.543
500	R	0.593	0.598	0.653	0.639	0.552	0.588	0.527	0.571	0.499	0.528
	ChatGPT	0.593	0.597	0.652	0.638	0.551	0.588	0.526	0.570	0.498	0.528
250	R	0.639	0.636	0.647	0.665	0.506	0.566	0.479	0.548	0.409	0.464
	ChatGPT	0.638	0.635	0.646	0.664	0.505	0.566	0.479	0.547	0.409	0.463

Tablo 5 incelendiğinde R ve ChatGPT ile kestirilen madde ayırt edicilik indekslerinin neredeyse eşit olduğu, bazılarında 0.001’lik farkın olduğu görülmektedir. Bu farkın programların yuvarlama mekanizmasından kaynaklandığı düşünülmektedir.

Tablo 6.
Madde ayırt edicilik indeksleri (K=20)

N	250		500		1000	
Madde	R	ChatGPT	R	ChatGPT	R	ChatGPT
1	0.525	0.525	0.519	0.519	0.528	0.528
2	0.568	0.568	0.568	0.568	0.571	0.571
3	0.549	0.549	0.573	0.573	0.592	0.592
4	0.564	0.564	0.549	0.549	0.599	0.599
5	0.429	0.429	0.491	0.491	0.557	0.557
6	0.475	0.475	0.515	0.515	0.559	0.559
7	0.519	0.519	0.535	0.535	0.594	0.594
8	0.462	0.462	0.522	0.522	0.610	0.610
9	0.524	0.524	0.582	0.582	0.614	0.614
10	0.588	0.588	0.630	0.630	0.647	0.647
11	0.584	0.584	0.624	0.624	0.625	0.625
12	0.537	0.537	0.532	0.532	0.604	0.604
13	0.626	0.626	0.639	0.639	0.685	0.685
14	0.452	0.452	0.497	0.497	0.587	0.587
15	0.575	0.575	0.609	0.609	0.664	0.664
16	0.350	0.350	0.453	0.453	0.538	0.538
17	0.532	0.532	0.550	0.550	0.561	0.561
18	0.328	0.328	0.474	0.474	0.479	0.479
19	0.356	0.356	0.454	0.454	0.523	0.523
20	0.290	0.290	0.285	0.285	0.429	0.429

Tablo 6 incelendiğinde tüm örneklem büyüklüğü ve madde sayısı koşullarında madde toplam test puanı korelasyon katsayıları bir diğer ifadeyle madde ayırt edicilik indeksleri ChatGPT ve R ile yapılan kestirimlerde eşit bulunmuştur. Buna ChatGPT’nin madde ayırt ediciliği kestiriminde R ile benzer performans sergilediği söylenebilir.

Güvenirliliğe ilişkin bulgular

Üretilen veriler bir ölçek verisi gibi ele alınarak ölçeğin bütüne ilişkin Cronbach alfa iç tutarlılık katsayısı hesaplanmıştır.

Tablo 7.

Cronbach Alfa kestirimleri

		Madde sayısı	
n		10	20
1000	R	0.85	0.91
	ChatGPT	0.85	0.91
500	R	0.86	0.90
	ChatGPT	0.86	0.90
250	R	0.88	0.88
	ChatGPT	0.88	0.88

Tablo 7 incelendiğinde tüm örneklem büyüklüğü ve madde sayısı koşullarında ChatGPT ve R ile elde edilen Cronbach alfa kestirimlerinin birbirine eşit olduğu görülmektedir. Bu durumda ChatGPT Cronbach alfa kestiriminde R ile benzer performans sergilemektedir.

TARTIŞMA, SONUÇ VE ÖNERİLER

Son yıllarda özellikle yazılım alanındaki gelişmelere paralel olarak veri biliminde de önemli adımlar atılmış ve gelişmeler kaydedilmiştir. Özellikle yapay zekanın pek çok alana uyarlanıp kullanılabildiği günümüzde veri bilimi ve istatistiksel analizler de payını almıştır. Yapay zekâ uygulamaları da bir takım istatistiksel analizleri yapabilmektedir ki doğru sonuçlar elde edilmesi ve etik kurallara uygun davranılması durumunda veri analizi için yapay zekadan faydalanmanın önünde herhangi bir engel yoktur. Bu noktada öncelikle emin olunması ve doğrulanması gereken durum yapay zekanın veri analizinde doğru bulgular sunup sunmadığıdır. Bu araştırmada sosyal bilimler alanında özellikle yapı geçerliğini belirlemek için sıklıkla başvurulmuş açımlayıcı faktör analizi, madde ayırt ediciliği ve güvenilirlik hesaplamalarının yapay zekâ ile yapılması durumunda elde edilecek bulgularının doğruluğunun incelenmesi amaçlanmıştır. Bu amaç doğrultusunda yapay olarak üretilen 250, 500 ve 1000 örneklem büyüklüğünde 10 ve 20 maddelik toplam 6 farklı koşuldaki veri setleri ChatGPT ve R programı kullanılarak analiz edilmiş ve sonuçları karşılaştırılmıştır. R programından elde edilen bulgular ChatGPT’den elde edilen bulguların ve dolayısıyla performansının değerlendirilmesinde referans değer olarak kullanılmıştır.

Araştırmada iç tutarlılık anlamında güvenilirlik belirleme yöntemi olan Cronbach alfa ve madde ayırt ediciliği hesaplamalarında R ve ChatGPT arasında fark gözlenmemiştir. Bu nedenle ChatGPT’nin Cronbach alfa ve madde ayırt ediciliği hesaplamalarında iyi bir performans sergilediği söylenebilir. Araştırma sonucunda R ve ChatGPT ile yapılan açımlayıcı faktör analizinde elde edilen KMO, Bartlett Küresellik testi sonucu, açıklanan varyans oranı ve K1 yöntemine göre belirlenen faktör sayısı bulgularının tüm örneklem ve madde sayısı koşullarında uyumlu olduğu sonucuna erişilmiştir. Buna göre Chat GPT verinin faktörleşmeye uygunluğunu kontrol etmede bir diğer ifadeyle açımlayıcı faktör analizinin varsayımlarını sınamada, geçerli faktör sayısını belirlemede ve açıklanan toplam varyans oranını belirlemede iyi bir performans sergilemektedir. Bu hesaplamalar ve kararlar bir ölçüt değere bağlı olarak verilmektedir (Izquierdo, OleaveAbad, 2014; Lloret, Ferreres, Hernandezve Tomas, 2017; Watkins, 2018). Örneğin faktör sayısı için özdeğerin 1’in üzerinde olması kuralı uygulanmıştır bu çalışmada ve kriterlere kurallara bağlı kararlarda ChatGPT iyi performans sergilemiştir. Bu nedenle araştırmacılar Chat GPT kullanarak, uygun komutlar ile açımlayıcı faktör analizini bu bulgular için kullanabilir. Elde edilecek bulgular benzer veri koşulları için doğru sonuçlar içerecektir.

Açımlayıcı faktör analizinin bir diğer önemli basamağı da madde faktör dağılımını belirlemek ve madde faktör yüklerini hesaplamaktır. Özellikle hangi maddenin hangi faktör altında yer alması gerektiğine karar vermek önemli bir adımdır. Varimax döndürme yapılarak R ve ChatGPT'nin madde faktör dağılımı bulguları incelendiğinde madde sayısı 10 olan veri setlerinde yalnızca 250 örneklem büyüklüğünde bir maddeye ilişkin farklı boyutlandırma yapıldığı görülmektedir. Diğer örneklem büyüklüklerinde her ikisi de tüm maddeleri aynı şekilde sınıflandırmıştır. Ancak madde sayısı 20 olan veri setlerinde uyumsuzluk sayısı artmıştır. Örneklem büyüklüğü arttıkça (500, 1000) daha fazla madde açısından farklı faktör sınıflandırmaları elde edildiği sonucuna erişilmiştir. R ile yapılan analizlerde maddelerin faktörlere verdiği yükler göz önünde bulundurularak araştırmacı kararı ile hangi maddenin hangi faktörde olacağına karar verilmiştir. ChatGPT ise yine maddelerin faktör yüklerini göz önünde bulundurarak hangi maddenin hangi boyutta olacağına kendi algoritması ile karar vermiştir. Burada önemli olan nokta R'da insan kararının ChatGPT'de ise yapay zekâ kararının söz konusu olmasıdır ve bu iki karar maddenin farklı faktörlere de önemli ölçüde yük vermesi durumunda farklı olmuştur. Bu durum göz önünde bulundurularak ChatGPT ile açımlayıcı faktör analizi yapılması durumunda madde faktör dağılımı belirlenirken mutlaka kontrol edilmelidir. Aksi durumunda hatalı denilebilecek madde faktör dağılımları elde edilebilir.

Normal şartlarda da herhangi bir istatistik paket program kullanılarak yapılacak açımlayıcı faktör analizinde madde faktör dağılımını belirlemede araştırmacı hem madde faktör yükünü hem de maddenin ölçtüğü düşünülen yapı ve boyutu göz önünde bulundurarak karar vermelidir. Erkuş (2011) bir istatistik programının ya da yazılımının kullanılabilirliği olmasının veri analizi için yeterli olmayacağını belirtmektedir. Bu araştırmada da belirli kriterlere göre karşılaştırma yapmayı gerektiren adımlarda ChatGPT iyi bir performans sergilemiştir ancak birkaç durumu göz önünde bulundurarak yorumlanması ve karar verilmesi gereken noktalarda performansı düşmüştür. Dolayısıyla ChatGPT'nin veri analizinin yorum gerektiren kararlarında henüz yetersiz olduğu söylenebilir. Özellikle binişik madde olması durumunda yani bir maddenin birkaç faktöre yük vermesi durumunda o maddenin daha dikkatli incelenerek hangi boyutta olacağı hatta bazı durumlarda ölçekten çıkartılması kararı verilmelidir (Braekenve Van Assen, 2017; Howard, 2016; Shrestha, 2021). Yapay zekanın binişik maddeler için hangi faktörde olması gerektiğine karar vermede performansının düşük olduğu sonucuna erişilmiştir. Bu nedenle araştırmacıların bu kararı bütünüyle yapay zekaya bırakmaması önerilebilir.

Bir veri analizi olarak ChatGPT'nin performansı bu araştırmada açımlayıcı faktör analizi, Cronbach alfa katsayısı ve madde ayırt ediciliği üzerinden incelenmiştir. Bu araştırmanın bulguları bu analizler için yönlendirici bilgiler sunabilir, diğer analizler için ise dolaylı bir öngörü sunacaktır. Yorumlama gerektirmeyen yalnızca hesaplama gerektiren ya da bir ölçüt değerle karşılaştırılarak karar verilmesini gerektiren analizlerde performansı iyi olabilir. Ancak bunun yalnızca bir öngörü olduğu unutulmamalıdır ve başka analizlerdeki performansı da incelenerek ilgili analizler için doğru bulgular sunup sunmadığı test edilmelidir. Herşeyin ötesinde sosyal bilimlerde özellikle bir yapı hakkında araştırma yapılırken kuramsal yapı hatta ilişkili olan diğer yapıların da göz önünde bulundurulması gerekmektedir. Ve mevcut koşullarda ChatGPT bunu yapamamaktadır bu nedenle duyuşsal, bilişsel ve davranışsal özellikler analiz edilirken insanın bütünüyle dışarıda bırakılması çok mümkün değildir.

KAYNAKÇA

- Arafat, S. M. Y., Chowdhury, H. R., Qusar, M. S., & Hafez, M. A. (2016). Cross-cultural adaptation and psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health*, 5(3), 129–136. <https://doi.org/10.5455/jbh.20160615121755>
- Arbağ, S. S., & Ertekin, E. (2020). Matematik eğitimi lisansüstü tezlerindeki geçerlik ve güvenilirlik çalışmalarının incelenmesi. *OPUS International Journal of Society Researches*, 16(31), 3924–3957.
- Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *SSRN*. <https://ssrn.com/abstract=4337484>
- Bandalos, D. L., & Finney, S. J. (2018). *Factor analysis: Exploratory and confirmatory*. The Guilford Press.
- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., & Skolits, G. J. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(18), 1–13.
- Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, 22(3), 450–466. <https://doi.org/10.1037/met0000074>
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2014). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. Pegem Akademi Yayıncılık.
- Dehouche, N. (2021). Harnessing the power of AI in education: ChatGPT as an intelligent tutoring system. *Journal of Artificial Intelligence in Education*, 31(3), 456–469.
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.
- Dooley, K. (2002). Simulation research methods. In J. Baum (Ed.), *Companion to organizations* (pp. 829–848). Blackwell.
- Erkuş, A. (2011). *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler*. Pegem Akademi.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Allyn & Bacon.
- Gillani, N., Eynon, R., Chiabaut, C., & Finkel, K. (2023). Unpacking the “black box” of AI in education. *Educational Technology & Society*, 26(1), 99–111.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Saunders.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). Exploratory factor analysis. In J. F. Hair, W. C. Black, B. J. Babin, & R. E. Anderson (Eds.), *Multivariate data analysis* (7th ed., pp. 88–127). Prentice Hall.
- Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1), 51–62. <https://doi.org/10.1080/10447318.2015.1087664>
- Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, 26(4), 395–400. <https://doi.org/10.7334/psicothema2013.349>
- Kartal, S. K., & Dirlik, E. M. (2016). Geçerlik kavramının tarihsel gelişimi ve güvenirlikte en çok tercih edilen yöntem: Cronbach alfa katsayısı. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(4), 1865–1879.
- Kasneci, E., Kasneci, G., Stede, M., & Naumann, F. (2023). ChatGPT’s role in automated text analysis: Strengths, limitations, and future directions. *Computational Linguistics*, 49(1), 99–117.
- Liang, X., Xiao, T., Liu, H., Zhou, W., Wang, Y., & Li, J. (2023). Transformer-based models and their impact on AI research: A review. *Artificial Intelligence Review*, 56(1), 57–88. <https://doi.org/10.1007/s10462-023-10277-0>
- Lloret, S., Ferreres, A., Hernandez, A., & Tomas, I. (2017). The exploratory factor analysis of items: Guided analysis based on empirical data and software. *Anales de Psicología*, 33(2), 417–432. <https://doi.org/10.6018/analesps.33.2.270211>
- Mhlanga, D. (2023). Open AI in education: The responsible and ethical use of ChatGPT towards lifelong learning. *SSRN*. <https://ssrn.com/abstract=4354422> <https://doi.org/10.2139/ssrn.4354422>

- Nemorin, S., Gillani, N., & Wood, S. (2023). AI and plagiarism: Challenges and solutions in academic integrity. *Journal of Educational Technology*, 32(4), 245–258.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS*. Open University Press/McGraw-Hill.
- R Development Core Team. (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Revelle, W., & Revelle, M. W. (2015). Package ‘psych’. *The Comprehensive R Archive Network*, 337(338), 161–165. <https://CRAN.R-project.org/package=psych>
- Rudolph, J., & Tan, S. C. (2023). ChatGPT and generative AI in academic settings: Friend or foe? *Journal of Educational Technology*, 24(3), 215–230. <https://doi.org/10.1007/s11423-023-10211-7>
- Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4–11. <https://doi.org/10.12691/ajams-9-1-2>
- Singh, V., Chen, S. S., Singhania, M., Nanavati, B., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries—A review and research agenda. *International Journal of Information Management Data Insights*, 2(2), 100094.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103.
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach’s alpha. *International Journal of Medical Education*, 2, 53–55.
- Terwiesch, C. (2023). GPT-4: The next generation of generative AI for business and academia. *Harvard Business Review*, 101(4), 45–50.
- Tlili, A., Huang, R., Gao, W., & Kinshuk. (2023). AI and education: From theory to practice. *Computers and Education: Artificial Intelligence*, 4, 100150. <https://doi.org/10.1016/j.caeai.2023.100150>
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practices. *Journal of Black Psychology*, 44(1), 36–65. <https://doi.org/10.1177/0095798418771807>