T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

FAKTÖR ANALİZİ ÇÖZÜMLEMESİNDE BİRLEŞİK ENTROPİ YAKLAŞIMI

Nida ORUÇ

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Dr. Öğr. Üyesi Doğan YILDIZ

Haziran, 2024

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

FAKTÖR ANALİZİ ÇÖZÜMLEMESİNDE BİRLEŞİK ENTROPİ YAKLAŞIMI

Nida ORUÇ tarafından hazırlanan tez çalışması 14.06.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS** olarak kabul edilmiştir.

Dr. Öğr. Üyesi Doğan YILDIZ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Dr. Öğr. Üyesi Doğan YILDIZ, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. Atıf EVREN, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Erhan USTAOĞLU, Üye
Marmara Üniversitesi _____

Danışmanım Dr. Öğr. Üyesi Doğan YILDIZ sorumluluğunda tarafımca hazırlanan "Faktör Analizi Çözümlemesine Birleşik Entropi Yaklaşımı" başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Nida ORUÇ

İmza



Aileme

TEŞEKKÜR

Bu tez çalışmasında değerli fikirleriyle her daim bana destekçi olan, Yüksek Lisans eğitimim boyunca engin bilgileri, tecrübelerini ve emeğini hiçbir zaman esirgemeyen, her zaman bana yol gösteren sayın danışman hocam Dr. Öğr. Üyesi Doğan YILDIZ'a sonsuz teşekkürlerimi sunarım.

Her zaman her konuda bana yardımcı olan, yönlendiren ve sabırla desteğini esirgemeyen Arş. Gör. Dr. Mehmet Şamil GÜNEŞ'e teşekkür ederim.

Çalışma hayatım ve tez sürecimde daima yanımda olan, maddi ve manevi desteklerini hiç esirgemeyen sevgili ablam Sena ORUÇ GÜNELİ'ye teşekkür ederim.

Tüm hayatım boyunca daima yanımda olan, sonsuz özveri ve emekleri ile bugünlere gelmemi sağlayan, beni her koşulda destekleyen, eğitim hayatımda en az benim kadar emek veren değerli annem Ahsen ORUÇ ve değerli babam Çetin Derviş ORUÇ'a teşekkürü bir borç bilirim.

Nida ORUÇ

İÇİNDEKİLER

TEŞEKKÜR iv
SİMGE LİSTESİ vii
KISALTMA LİSTESİ viii
ŞEKİL LİSTESİ ix
TABLO LİSTESİ xi
ÖZET xii
ABSTRACT xiv
1 GİRİŞ **16**
1.1 Literatür Özeti 16
1.2 Tezin Amacı 3
1.3 Hipotez 3
2 TEORİK ÇERÇEVE **5**
2.1 Bibliyometrik Analiz 5
2.2 Entropi 15
2.2.1 Kavramsal Temeller 15
2.2.2 Entropinin Sınıflandırılması 16
2.2.3 İstatistiksel Belirsizliğin Ölçülmesi 18
2.2.4 Bilgi Teorisinde Entropi 19
2.2.5 Entropinin Özellikleri 22
2.2.6 Shannon Entropisi 23
2.2.7 Ortak Entropi 28
2.2.8 Ortak Bilgi 29
2.3 Faktör Analizi 30
2.3.1 Kavramsal Temeller 31
2.3.2 Faktör Analizi Çeşitleri 33
2.3.3 Faktör Analizi Varsayımları 34

2.3.4 Faktör Analizi Aşamaları 35
2.3.5 Verinin Uygunluğu 36
2.3.6 Faktörlerin Belirlenmesi 38
2.3.7 Faktör Sayısının Belirlenmesi 40
2.3.8 Faktörin İsimlendirilmesi 41
2.3.9 Faktör Döndürmesi 42
2.3.10 Faktör Skorları 45
2.3.11 Faktörin Yorumlanması 46
2.3.12 Faktör Analizi Uygulamalarında Karşılaşılan Sorunlar 47
2.4 Faktör Analizinin Ortak Entropide Uygulanması 47
3 UYGULAMA 52
3.1 Amaç 52
3.2 Veri Seti 52
3.3 Faktör Analizini Uygulaması 63
3.4 Ortak Entropi Uygulaması 66
4 SONUÇ 71
4.1 Sonuç ve Öneriler 71
KAYNAKÇA **72**
TEZDEN ÜRETİLMİŞ YAYINLAR **76**

SİMGE LİSTESİ

H	Entropi

KISALTMA LİSTESİ

AFA	Açıklayıcı Faktör Analizi
DFA	Doğrulayıcı Faktör Analizi
KMO	Kaiser-Meyer-Olkin
TBA	Temel Bileşen Analizi
GSYİH	Kişi Başına Gayri Safi Yurtiçi Hasıla

ŞEKİL LİSTESİ

Şekil 2.1 Faktör analizinin yıllara göre bilimsel üretim sayısı **6**

Şekil 2.2 Faktör analizinde en ilgili yazarlar **6**

Şekil 2.3 Faktör analizinde yazarların zaman içindeki üretimi **7**

Şekil 2.4 Faktör analizi kavramlarında en sık geçen kelimelere ait kelime bulutu . **7**

Şekil 2.5 Entropi ve ortak entropi kavramlarının yıllara göre bilimsel üretim sayısı **8**

Şekil 2.6 Entropi ve ortak entropi kavramlarına ilişkin dergi, yazar ve anahtar kelime ilişkisi **9**

Şekil 2.7 Entropi ve ortak entropi kavramlarında en ilgili yazarlar **9**

Şekil 2.8 Entropi ve ortak entropi kavramlarında yazarların zaman içindeki üretimi **10**

Şekil 2.9 Entropi ve ortak entropi kavramlarında en sık geçen kelimeler **10**

Şekil 2.10 Entropi ve ortak entropi kavramlarında en sık geçen kelimelere ait kelime bulutu **11**

Şekil 2.11 Faktör analizi ve ortak entropi kavramlarının yıllara göre bilimsel üretim sayısı **12**

Şekil 2.12 Faktör analizi ve ortak entropi kavramlarına ilişkin dergi, yazar ve anahtar kelime ilişkisi **12**

Şekil 2.13 Faktör Analizi ve ortak entropi kavramlarında en ilgili yazarlar **13**

Şekil 2.14 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimeler . **13**

Şekil 2.15 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimelere ait kelime bulutu **14**

Şekil 2.16 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimelerin sayısı **14**

Şekil 2.17 Çambel tarafından yapılan entropi sınıflandırması **16**

Şekil 2.18 Shannon'nun tanımladığı iletişim sistemi **24**

Şekil 2.19 Karşılıklı bilgi ve entropi arasındaki ilişki **30**

Şekil 2.20 Faktör analizinin görsel temsili **31**

Şekil 2.21 Birleşik entropi matrisinin görselleştirilmiş hali **51**

Şekil 3.1 Çocuk ölümünün en az olduğu 15 ülke **54**

Şekil 3.2 Çocuk ölümünün en çok olduğu 15 ülke **55**

Şekil 3.3 İhracat oranının en az olduğu 15 ülke **55**

Şekil 3.4 İhracat oranının en çok olduğu 15 ülke **56**

Şekil 3.5 Sağlık harcamaları oranının en az olduğu 15 ülke **56**

Şekil 3.6 Sağlık harcamaları oranının en çok olduğu 15 ülke **57**

Şekil 3.7 İthalat oranının en az olduğu 15 ülke **57**
Şekil 3.8 İthalat oranının en çok olduğu 15 ülke **58**
Şekil 3.9 Gelirin en az olduğu 15 ülke **58**
Şekil 3.10 Gelirin en çok olduğu 15 ülke **59**
Şekil 3.11 Enflasyonun en düşük olduğu 15 ülke **59**
Şekil 3.12 Enflasyonun en yüksek olduğu 15 ülke **60**
Şekil 3.13 Beklenen yaşam süresi en düşük olduğu 15 ülke **60**
Şekil 3.14 Beklenen yaşam süresi en yüksek olduğu 15 ülke **61**
Şekil 3.15 Doğruganlık düzeyinin en düşük olduğu 15 ülke **61**
Şekil 3.16 Doğurganlık düzeyinin en yüksek olduğu 15 ülke **62**
Şekil 3.17 Kişi başına düşen gayri safi yurtiçi hasılat düzeyinin en düşük olduğu 15 ülke **62**
Şekil 3.18 Kişi başına düşen gayri safi yurtiçi hasılat düzeyinin en yüksek olduğu 15 ülke **63**
Şekil 3.19 Scree plot **64**
Şekil 3.20 Açıklanan toplam varyans **65**
Şekil 3.21 Faktör Analizi ve Ortak Entropi yöntemlerinden elde edilen faktörlerin karşılaştırılması **70**

TABLO LİSTESİ

Tablo 2.1: Zar atma ile Para atma deneylerinin kıyaslanması **28**

Tablo 2.2 Veri setinde kullanılan değişkenler..... **49**

Tablo 3.1 Veri setinde kullanılan değişkenler..... **53**

Tablo 3.2 Kaiser-Meyer-Olkin ve Bartlett Testi **63**

Tablo 3.3 Toplam Varyans..... **64**

Tablo 3.4 SPSS ile faktör analizinden elde edilen faktörler **65**

Tablo 3. 5 Değişkenlere ait entropi değerleri..... **66**

Tablo 3. 6 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar..... **66**

Tablo 3. 6 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar(devamı) **67**

Tablo 3. 6 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar(devamı) **68**

Tablo 3. 7 Birleşik Entropi Matrisi **68**

Tablo 3. 8 Entropi- Korelasyon Matrisi **68**

Tablo 3. 8 Entropi- Korelasyon Matrisi (devamı)..... **69**

Tablo 3. 9 Ortak entropi ile elde edilen faktörler..... **69**

ÖZET

FAKTÖR ANALİZİ ÇÖZÜMLEMESİNDE BİRLEŞİK ENTROPİ YAKLAŞIMI

Nida ORUÇ

İstatistik Anabilim Dalı

İstatistik Programı

Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Doğan YILDIZ

Faktör analizi, belirli bir olguyla ilgili ve birbirleriyle ilişkili çok sayıda değişkenin birbirleriyle olan ilişkisini inceleyerek, değişkenlerin sayısını azaltmak ve ilişkisiz yeni değişkenler yaratmak için kullanılan çok değişkenli istatistiksel bir analizdir. Tezimizde farklı bir yaklaşım olarak, faktör çıkarmada birleşik entropiden yararlanılmıştır. Buradan elde edilen "Faktörleşme" ile "Açıklayıcı Faktör Analizi" sonuçları karşılaştırılarak benzerlik ve farklılıkları tartışılmıştır.

Tez kapsamında Kaggle sitesinden "Unsupervised Learning on Country Data" başlıklı veri seti alınarak uygulama yapışmıştır. Veri setinin ülkeler bazında genel gelişimini belirleyen sosyo-ekonomik ve sağlık faktörtörlerini kullanarak katagorize edilmesi amaçlanmıştır. Bu bağlamda 167 ülke üzerinden bir değerlendirme söz konusudur. Uygulama kısmında SPSS, Python ve excel kullanılarak sonuçlar değerlendirilmiştir.

Bu tez çalışmasında, istatistik ile entropiyi birleştiren bir model kullanılmıştır. Çalışmada, ilk olarak bibliyometrik analiz yapışmıştır. Uygulama kısmında ise ilk olarak istatistikte yaygın olarak kullanılan faktör analizi sonuçları elde edilmiştir.

Ardında ise kovaryans matrisi yerine birleşik entropi matrisi kullandığımız entropi tabanlı yaklaşımla elde edilen matris ile bir karşılaştırma yapılmıştır. Sonrasında ise birleşik entropi matrisi ile elde edilen sonuçların boyut indirgemenin sağlanıp sağlanmadığını değerlendirilmiştir. Bu bağlamda birleşik entropi matrisi ile elde ettiğimiz sonuçlar ile faktör analizi ile elde ettiğimiz sonuçlar karşılaştırılmıştır. Son durumda ise birleşik entropi matrisi ile elde edilen sonuçların boyut indirgemenin sağlanıp sağlanmadığını ve avantajları ile dezavantajları üzerinde durulmuş olup sonuçlar gösterilmiştir. Ek olarak, bu çalışma ile entropiye teorik destek ile aslında varyans-kovaryans ölçülerinin kullanım alanlarına uygulanabileceği ve desteklenebileceğini göstermek hedeflenmiş ve çözülmesi gereken bazı sorunlara rağmen, alternatif olarak bilinen yöntemlerin yerine önerilmektedir.

Anahtar Kelimeler: Bibliyometrik analiz, faktör analizi, entropi, bilgi teorisi, ortak entropi.

ABSTRACT

JOINT ENTROPY APPROACH TO FACTOR ANALYSIS SOLUTION

Nida ORUÇ

Department of Statistics

Master of Science Thesis

Supervisor: Dr. Öğr. Üyesi Doğan YILDIZ

Factor analysis is a multivariate statistical analysis used to reduce the number of variables and create new unrelated variables by examining the relationship between many interrelated variables related to a particular phenomenon. As a different approach in our thesis, joint entropy was used in factor extraction. The "Factorization" and "Explanatory Factor Analysis" results obtained here were compared and their similarities and differences were discussed.

Within the scope of the thesis, the data set titled "Unsupervised Learning on Country Data" was taken from the Kaggle site and applied. It is aimed to categorize the data set using socio-economic and health factors that determine the general development of the data set on a country basis. In this context, there is an evaluation over 167 countries. In the application part, the results were evaluated using SPSS, Python and Excel.

In this thesis study, a model combining statistics and entropy was used. In the study, bibliometric analysis was used first. In the application part, firstly, factor analysis results, which are widely used in statistics, were obtained. Then, a comparison was made with the matrix obtained with the entropy-based approach, in which we used the joint entropy matrix instead of the covariance matrix. Afterwards, it was evaluated whether the results obtained with the combined entropy matrix achieved

dimensionality reduction. In this context, the results we obtained with the combined entropy matrix were compared with the results we obtained with factor analysis. In the last case, it is emphasized whether the dimension reduction of the results obtained with the combined entropy matrix is achieved and its advantages and disadvantages, and the results are shown. In addition, this study aims to show that, with theoretical support for entropy, variance-covariance measures can actually be applied and supported in their areas of use, and despite some problems that need to be solved, it is recommended instead of known alternative methods.

Keywords: Bibliometrics analysis, factor analysis, entropy, information theory, joint entropy.

1
GİRİŞ

1.1 Literatür Özeti

Ömürbek vd. (2021) aktardığına göre, entropi çok çeşitli bilim alanlarında kullanılan bir yöntem olmakla birlikte geniş bir uygulama alanına sahiptir. Bu kavramın çok çeşitli uygulamaları nedeniyle, temel çerçevesinden ayrılmadan her alanda kendisine özgü farklı bir formülasyonu mevcuttur [1].

Altaylıgil (2008)'in aktardığına göre, 1822-1888 yılları arasında yaşamış olan Alman fizikçi Rudolf Clausius ilk kez entropi terimini kullandı. 1865 yılında termodinamik bir sistemin enerji miktarının bir ölçüsü olarak entropiyi tanımlamıştır ve Clausius'un "Mechanical Theory of Heat" kitabında bu kavram üzerinde durulmuştur [2].

Çetinkaya (1999)'nın aktardığına göre, enerjinin korunumu prensibi termodinamiğin ilk yasasıdır. Bu temel prensip, bir sistemdeki enerjinin yoktan var edilemeyeceğini veya varken yok edilemeyeceğini ve ek olarak sadece bir şekilde dönüştürülebileceğini ifade etmektedir. Termodinamiğin ikinci yasasında ise, enerjinin bir şekilden başka bir şekle dönüşmesi işleminin aşamasında belirli bir yönde gerçekleşeceğini ve ters yönde gerçekleşmesinin mümkün olmadığını söylemektedir [3].

Taslaman (2006)'ın aktardığına göre, teoloji alanında da entropi kavramından yararlanılmıştır. Termodinamiğin ikinci yasasına göre, evrendeki düzensizliğin sürekli olarak artmaktadır ve bu artış tek yönlü bir eğilim göstermektedir. Bu yasa, ateistler ve ateistler arasında evrenin başlangıcı ve sonu olup olmadığına dair tarihsel tartışmalar için çok önemlidir [4].

Altaylıgil (2008)'in aktardığına göre, entropi kavramı birçok farklı alanlarda geniş çerçevede kullanılmasın yanı sıra, istatistikte de kullanılmaktadır. Ludwig Boltzmann 1877 yılında Clausius'un geliştirmiş olduğu entropi kavramını kullanarak gaz moleküllerinin dağılımındaki belirsizliği ölçmek için benzer bir formülü istatistik alanında geliştirmiştir. Bu bağlamda 1877 yılında Ludwig

Boltzmann olasılık ile entropi kavramı arasındaki ilişkiyi keşfetti. Böylece entropiyi istatistiksel olarak tarihte ilk kez ortaya koydu [2].

Baray (2003)'ın aktardığına göre, olasılık ve entropiyi bağlayan Boltzmann'ın entropi formülünü ilerleyen senelerde iletişim teorisinde de yer edinmektedir ve Shannon tarafından ise veri transferi esnasında meydana gelen bilgi kaybının belirsizliğini ölçmek için kullanılmıştır [5].

Kullback (1997)'in aktardığına göre, bilgi teorisinde bu kavram Shannon entropisi olarak da bilinir. Shannon'ın görüşlerinden önce, bilgi genellikle belirsiz bir sisli akış olarak görülüyordu. Ancak Shannon ile birlikte bilginin iyi tanımlanmış ve en önemlisi ölçülebilir bir nicelik olduğunu gösterdi [6].

Darton (1980) ve Harman (1976)'ın aktardığına göre faktör analizi, Charles Spearman, Kelley, Holzinger, Thurstone, Karl Pearson, Thomson ve Burt, tarafında 20. yüzyılın ilk çeyreğinde yaptıkları çalışmalarla faktör analizi fikrini ortaya atmış ve geliştirmişlerdir. Bu konudaki temel eksen faktörünü Karl Pearson (1901) bulmuştur. Spearman (1904) ise, tek faktörlü ve iki faktörlü modelleri ortaya koymuştur. Garnett (1919) ise çok faktörlü modelinin incelemesini yapmış olup Thurstone (1931) ise bu çok faktörlü modeli geliştirmiş ve faktör döndürme terimini geliştirmiştir. Thurstone'un 1935'te yazmış olduğu "The Vector of Mind" isimli kitabında ise bu araştırmaları özetlemiştir [7].

Özdamar ve Dinçer (1987)'in aktardığına göre faktör analizi, birbirleri ile ilişkili değişkenleri ortak etkili olan ve faktör olarak bilinen ortogonal yapıdaki değişkenlere dönüştürmek ve gerektiği durumlarda boyut indirgemesi olarak adlandırılan bir süreç olduğunu belirtmektedir [8].

Daniel (1988)'in aktardığına göre faktör analizi, bir değişken grubu arasındaki kovaryans yapısının detaylı bir biçimde incelenmesi ve bu değişkenler arasındaki ilişkileri daha az sayıda gözlemlenebilir gizli değişkenler olarak adlandırılan faktörler aracılığıyla açıklamayı tasarlayan bir tekniktir [9].

Stapleton (1997)'un aktardığına göre faktör analizi, esas değişkenler arasındaki kovaryans yapısını detaylı bir şekilde incelenmesi ve bu değişkenler arasındaki ilişkileri, doğrudan gözlenemeyen ve faktör olarak isimlendirilen değişkenler aracılığıyla incelemeyi amaçlayan bir yöntemdir. Rennie (1997)'e göre ise faktör analizi, gözlenen değişkenlerin arasındaki ilişkileri esas alarak hesaplama

yöntemine dayalı analitik bir metodoloji olarak tanımlar. Bu yaklaşım, maksimum varyansı açıklayabilecek az sayıda açıklayıcı faktör elde etmeyi amaçlamaktır [9].

Tatlıdil (2002)'in aktardığına göre faktör analizi, çok değişkenli analiz yöntemlerinden biri olup temel olarak, p boyutlu bir uzayda birbirleriyle bağlantılı olan değişkenlerin bir araya getirilmesiyle daha az sayıda yeni değişkenlerin ortaya çıkarıldığı ve bu bağlamda çıkan yeni değişkenlerinde birbirinden bağlantısız olan analitik bir yöntemdir [10].

Albayrak (2005)'ın aktardığına göre faktör analizi, çok karmaşık analizle karşılaşıldığı zamanlarda, Kanonik Korelasyon Analizi, Kümeleme Analizi ve Çok Boyutlu Ölçekleme Analizi gibi kullanılabilecek bir yöntemdir [11].

1.2 Tezin Amacı

Bu tez kapsamında ilk olarak bibliyometrik analiz yapılmıştır. Ardından ise bu çalışmanın temel hedefi olan, belirli bir veri seti ile ilk olarak faktör analizi gerçekleştirilmesidir ve bu analizin hemen ardından ise aynı veri seti kullanılarak faktör analizinin başındaki gözlemlenen değişkenlerin arasındaki bağımlılığın ölçülmesinde kullanılan korelasyon matrisi yerine entropik tabanlı yöntem kullanılacaktır. Değişkenlerin entropi değerlerinden faydalanılarak birleşik entropi matrisi oluşturulacak ve elde edilen bu matrisi faktör analizindeki varyans-kovaryans matrisi yerine kullanarak bu doğrultuda bu matris yardımı ile boyut indirgeme yönteminin etkinliği incelenecektir. Son aşamada, elde edilen sonuçların karşılaştırılması yapılacak olup bu doğrultuda farklı yaklaşımların avantajları ve dezavantajları incelenecektir.

1.3 Hipotez

Bu tez çalışması, kullanılan verinin faktör analizi ve entropi temelli yaklaşımın analizinin sonuçlarının karşılaştırılmasını değerlendirmeyi hedeflenmiştir. Genel anlamda tez çalışmasının temel hipotezlerinin sıralanmasını şu biçimde yapabiliriz. Birinci olarak, bibliyometrik analiz yapılacak ve kullandığımız veri ile literatürde oldukça yaygın olarak kullanılan faktör analizi uygulaması gerçekleştirilecektir. Bunu takip eden süreçte faktör analizindeki aynı amaç için entropi tabanlı yaklaşım kullanılarak birleşik entropi matrisi elde edilecek ve faktörler bulunacaktır. Bu iki

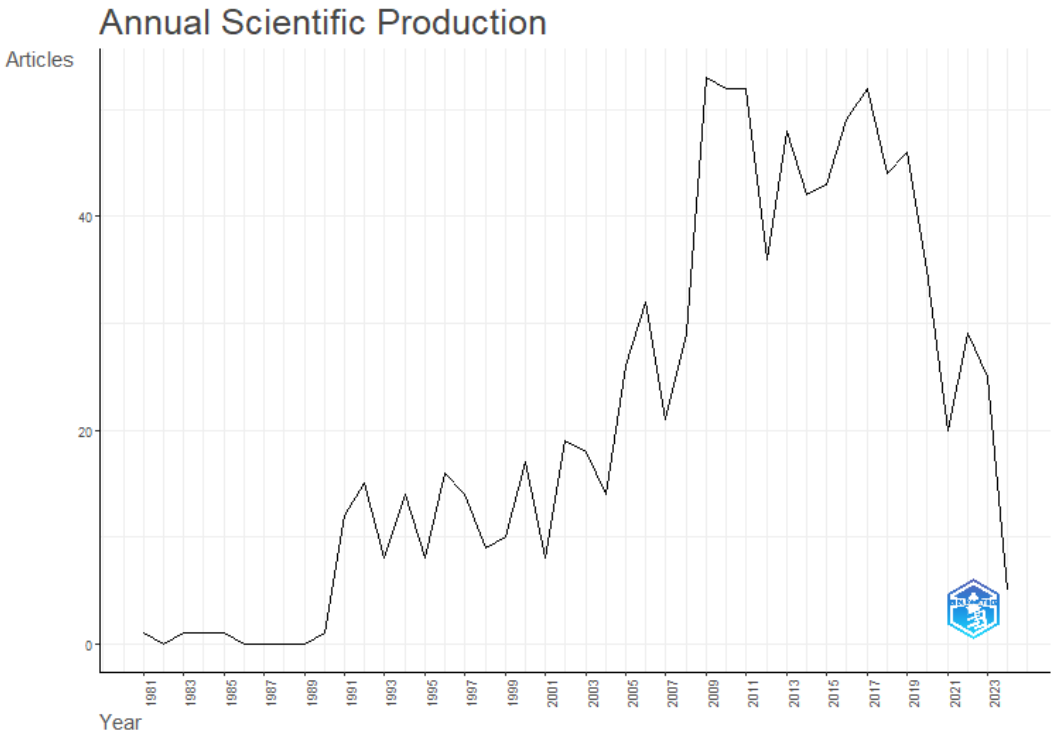
sonuç karşılaştırmalı olarak ölçülecektir. İkinci olarak ise kovaryans matrisi yerine birleşik entropi matrisi kullandığımız entropi tabanlı yaklaşımla elde edilen matris yardımı ile elde edilen sonuçların avantajları ve dezavantajları üzerinde durulacak ve değerlendirilecektir. Bu tez çalışmasının çıktıları ise entropinin istatistiksel anlamda daha sık kullanılmasına veya yapılacak olan araştırmalarda farklı bir yöntemin de kullanılabilecek olmasında yardımcı olması beklenebilir.

2
TEORİK ÇERÇEVE

2.1 Bibliyometrik Analiz

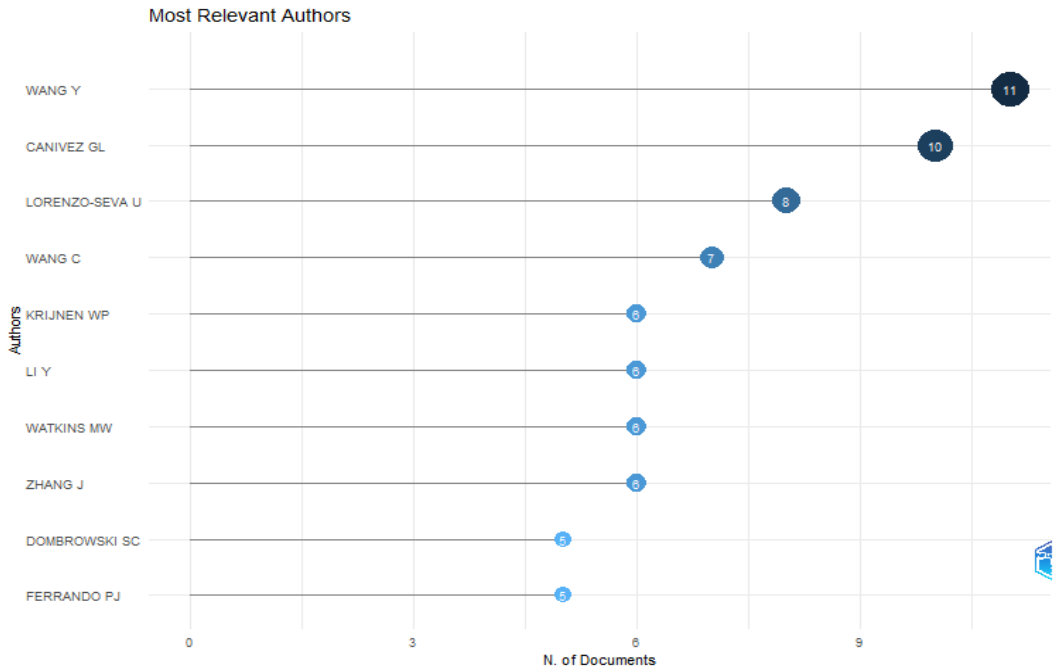
Bu bölüm faktör analizi, entropi ve ortak entropi ile ilgili makalelerin yazarları, konuları ve atıfları açısından belirlenen ağ yapılarını görsel olarak analizini elde etmeyi amaçlamaktadır. Bu nedenle, Web of Science (WOS) veri tabanında uluslararası dizinde öncelikle "factor analysis" anahtar kelimesi ile taranan makale çalışmaları gözlem birimi olarak kabul edildi ve bu makalelere ait bilgiler derlenerek veri seti oluşturuldu. Bu veri seti için en ilgili olan ilk 1000 veri alınmıştır. Ardından ise "entropy" ve "joint entropy" anahtar kelimeleri ile taranan makale çalışmaları gözlem birimi olarak kabul edildi ve bu çalışmalardan toplanan veriler derlenerek veri seti oluşturuldu. Son olarak, sadece "factor analaysis" ve "joint entropy" anahtar kelimelerini içeren makale çalışmaları gözlem birimi olarak kabul edildi ve bu makalelerden toplanan veriler toplanarak veri seti oluşturuldu. Bu çalışmada bibliyometrik analiz yapılmış ve R programları kullanılması tercih edilmiştir.

İlk olarak, "factor analysis" anahtar kelimesi üzerine yazılan makalelerin arasındaki ortak yazarlık durumu ve makalelerin konu bazındaki dağılımları belirlenmiştir. 1000 adet veri çekilmiş olup akabinde ise birkaç filtreleme ile 926 dökümana indirgenmiştir. Bu bağlamda 1981 ile 2024 yılları arasındaki verilere ulaşılmıştır. Bu çerçevede ortalama yıllık büyüme oranı ise %3.81 oranındadır. Dökümanlarda 2479 yazar olup bunlardan 154 tanesi tek yazarlıdır. Genel çerçeveye bakıldıktan sonra ise daha detaylı incelemek için öncelikle yıl bazlı bakmak gerekirse yıllık bilimsel üretim verileri Şekil 2.1'de gösterildiği gibidir.



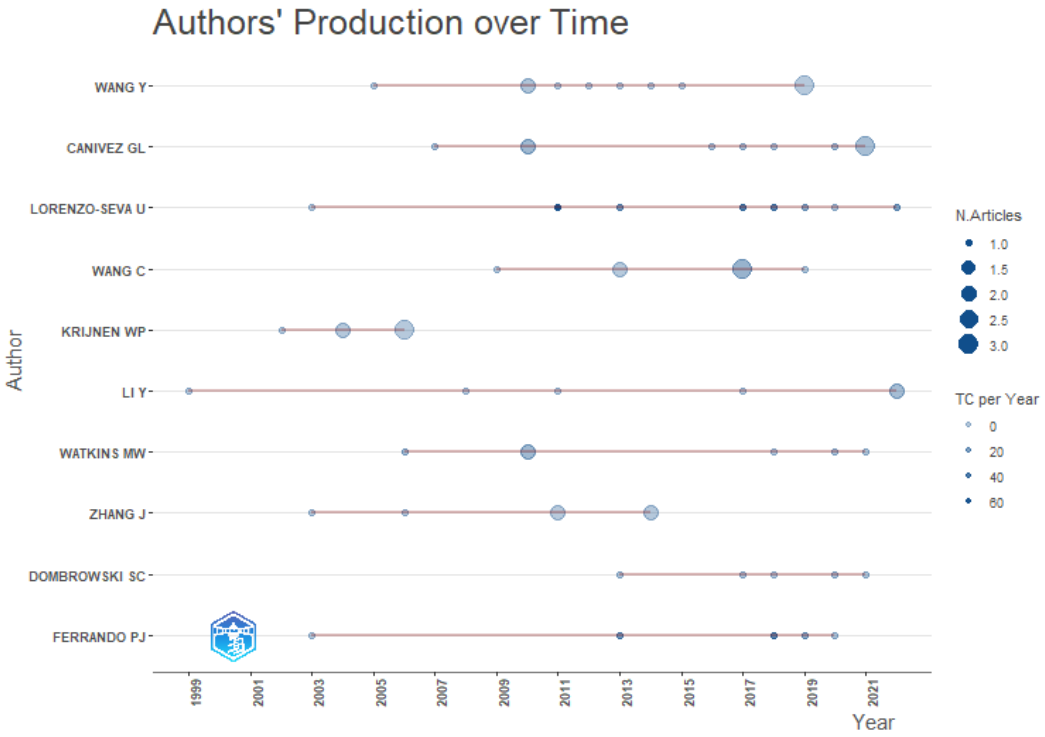
Şekil 2.1 Faktör analizinin yıllara göre bilimsel üretim sayısı

Bu çerçevede yıllar bazında bakıldığında en çok bilimsel üretim 59 tane ile 2009 yılı olurken, ardından 52 adet ile 2010 ve 2017 yılı gelmektedir. 2023 yılında ise 25 adet ve 2024 yılları içerisinde ise henüz 5 adet bilimsel üretim bulunmaktadır.



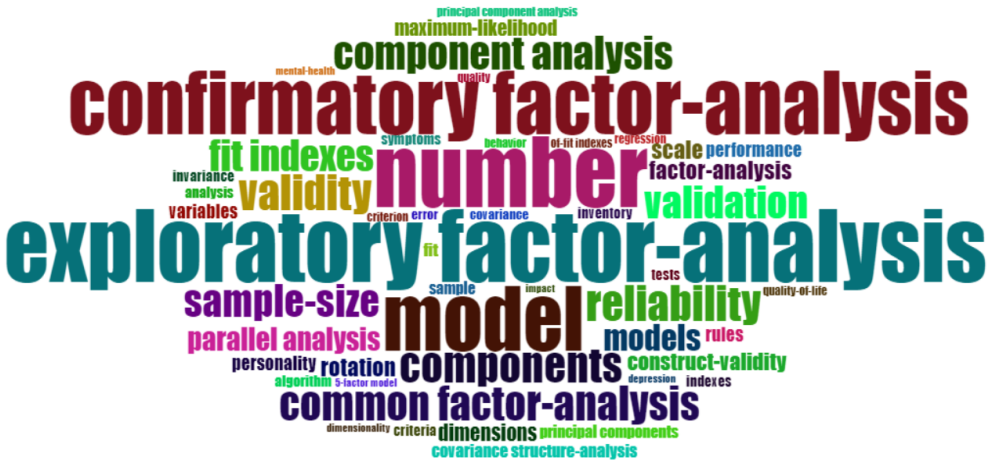
Şekil 2.2 Faktör analizinde en ilgili yazarlar

Şekil 2.2 incelendiğinde ise faktör analizi kavramı konusuna ait en ilgili yazarlar listesi bulunmaktadır. Buradaki yazarların zaman içindeki üretimi ise Şekil 2.3'te gösterilmektedir.



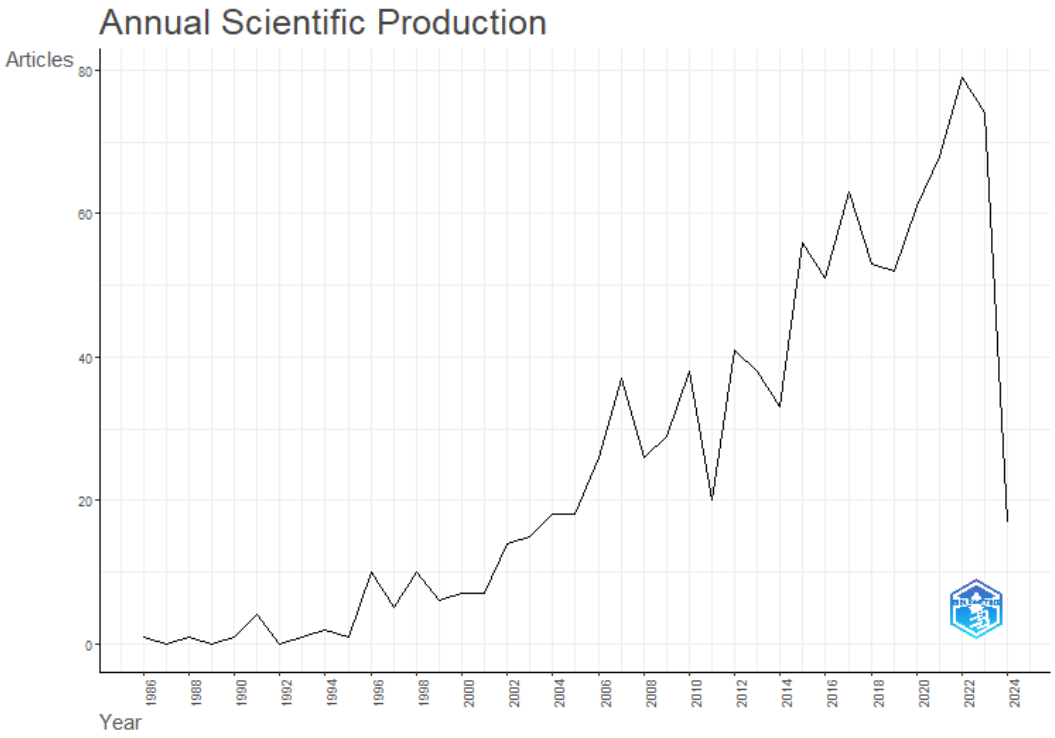
Şekil 2.3 Faktör analizinde yazarların zaman içindeki üretimi

Şekil 2.4'te ise kelime bulutu gözükmektedir. Bu kısımda anlaşılan ise dökümanlarda en sık tekrar eden kelimelerin kelime bulutu ile görselleştirilmiş şeklidir.



Şekil 2.4 Faktör analizi kavramlarında en sık geçen kelimelere ait kelime bulutu

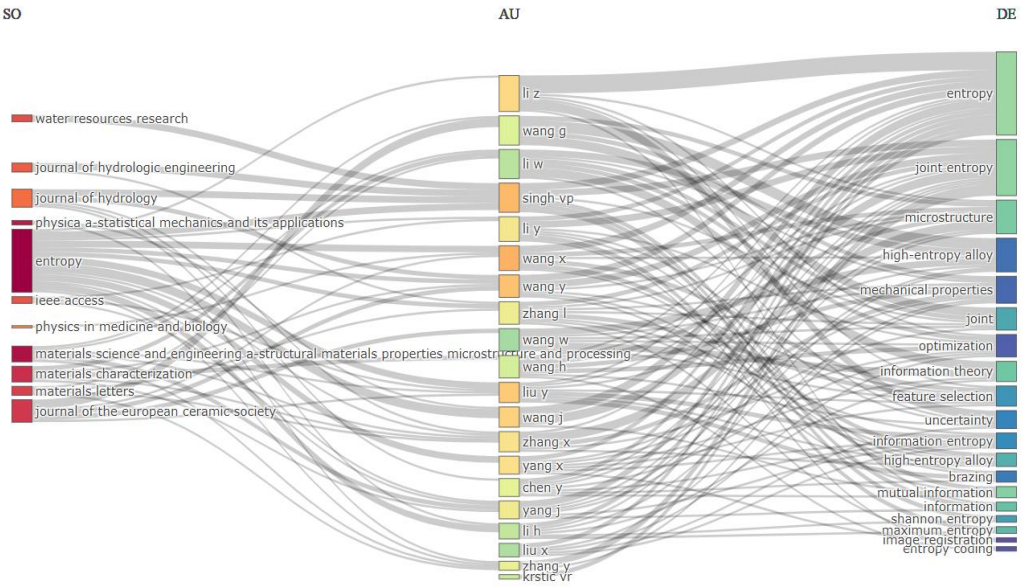
İkinci olarak, "entropy" ve "joint entropy" anahtar kelimesi üzerine yazılan makalelerin arasındaki ortak yazarlık durumu ve makalelerin konu bazındaki dağılımları belirlenmiştir. 1000 adet veri çekilmiş olup akabinde ise birkaç filtreleme ile 983 dökümana indirgenmiştir. Bu bağlamda 1986 ile 2024 yılları arasındaki verilere ulaşılmıştır. Bu çerçevede ortalama yıllık büyüme oranı ise %7.74 oranındadır. Dokümanlarda 2460 yazar olup bunlardan 96 tanesi tek yazarlıdır. Genel çerçeveye bakıldıktan sonra ise daha detaylı incelemek için öncelikle yıl bazlı bakmak gerekirse yıllık bilimsel üretim verileri Şekil 2.5'te gösterildiği gibidir.



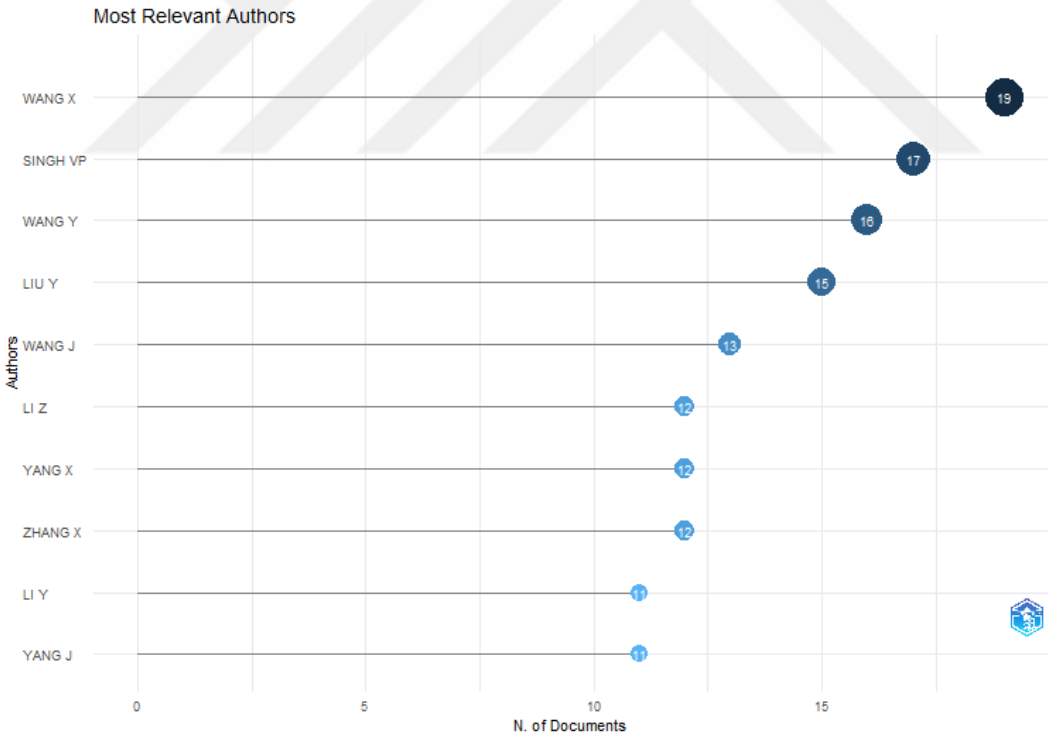
Şekil 2.5 Entropi ve ortak entropi kavramlarının yıllara göre bilimsel üretim sayısı

Bu çerçevede yıllar bazında bakıldığında en çok bilimsel üretim 79 tane ile 2022 yılı olurken, ardından 74 adet ile 2023 yılı gelmektedir. 2024 yılları içerisinde ise henüz 17 adet bilimsel üretim bulunmaktadır.

Şekil 2.6'da ise incelediğimiz kısım alanındaki etkin yazarların hangi dergilerde hangi anahtar kelimelerle kullandığının incelenmesini gösterilmektedir. Buradaki her kutucuğun farklı olmasındaki sebep ise o temsiliyet alanındaki en önemlilerini göstermesinden kaynaklanmaktadır.

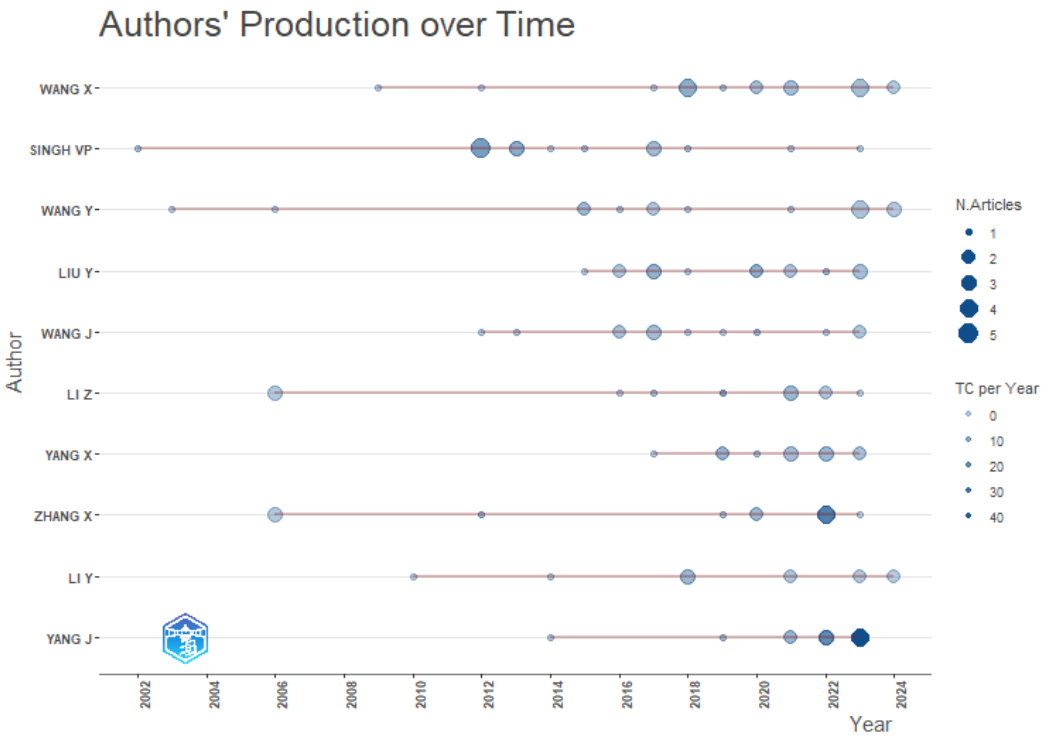


Şekil 2.6 Entropi ve ortak entropi kavramlarına ilişkin dergi, yazar ve anahtar kelime ilişkisi

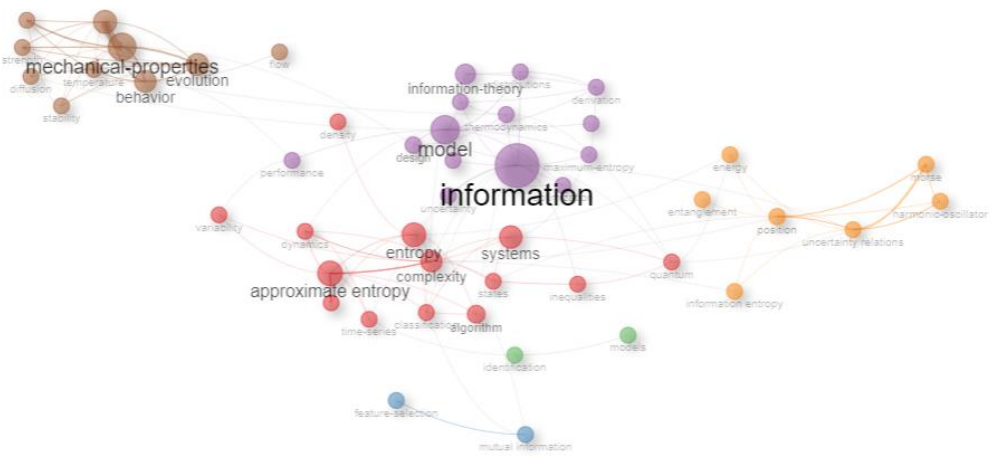


Şekil 2.7 Entropi ve ortak entropi kavramlarında en ilgili yazarlar

Şekil 2.7 incelendiğinde ise Entropi ve ortak entropi kavramlarında konusuna ait en ilgili yazarlar listesi bulunmaktadır. Buradaki yazarların zaman içindeki üretimi ise Şekil 2.8'de gösterilmektedir.

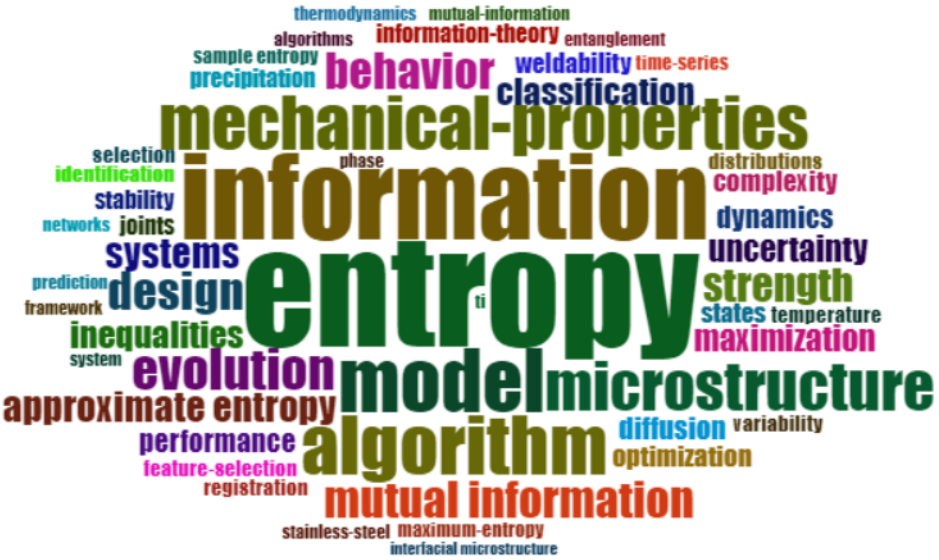


Şekil 2.8 Entropi ve ortak entropi kavramlarında yazarların zaman içindeki üretimi



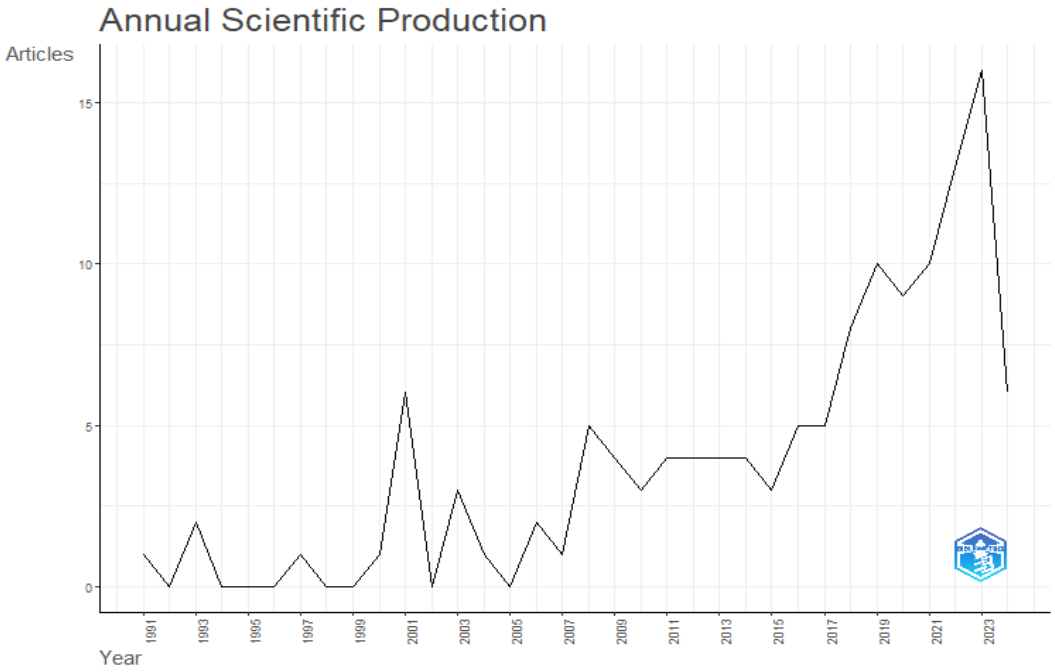
Şekil 2.9 Entropi ve ortak entropi kavramlarında en sık geçen kelimeler

Şekil 2.9'da ise "entropy" ve "joint entropy" kavramlarının geçtiği yayınlarda, en sık geçen kelimelerin kümelenmesi gözükmektedir. Şekil 2.10'da ise kelime bulutu gözükmektedir. Bu kısımda anlaşılan ise dökümanlarda en sık tekrar eden kelimelerin kelime bulutu ile görselleştirilmiş şeklidir.



Şekil 2.10 Entropi ve ortak entropi kavramlarında en sık geçen kelimelere ait kelime bulutu

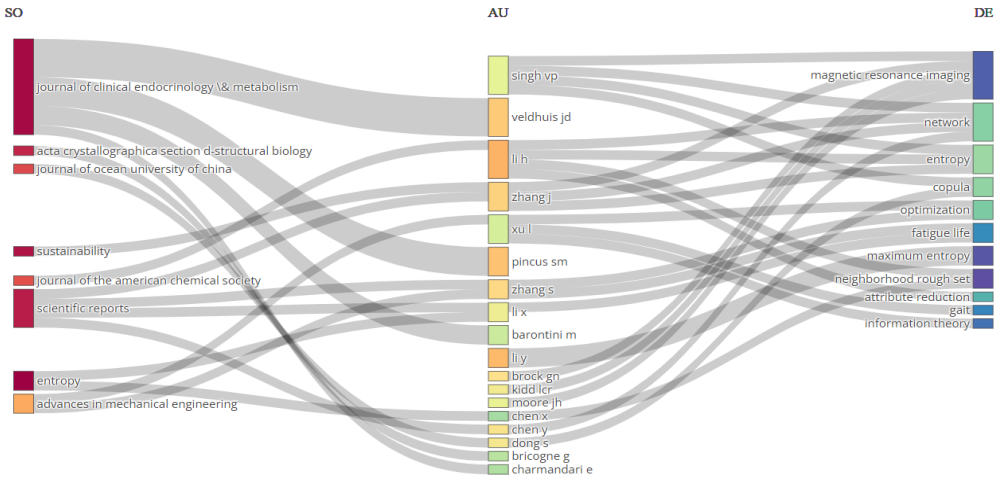
Son olarak ise "factor analysis" ve "joint entropy" anahtar kelimesi üzerine yazılan makalelerin arasındaki ortak yazarlık durumu ve makalelerin konu bazındaki dağılımları belirlenmiştir. 133 adet veri çekilmiş olup akabinde ise birkaç filtreleme ile 131 dökümana indirgenmiştir. Bu bağlamda 1991 ile 2024 yılları arasındaki verilere ulaşılmıştır. Bu çerçevede ortalama yıllık büyüme oranı ise %5.58 oranındadır. Dokümanlarda 532 yazar olup bunlardan 7 tanesi tek yazarlıdır. Genel çerçeveye bakıldıktan sonra ise daha detaylı incelemek için öncelikle yıl bazlı bakmak gerekirse yıllık bilimsel üretim verileri Şekil 2.11'de gösterildiği gibidir.



Şekil 2.11 Faktör analizi ve ortak entropi kavramlarının yıllara göre bilimsel üretim sayısı

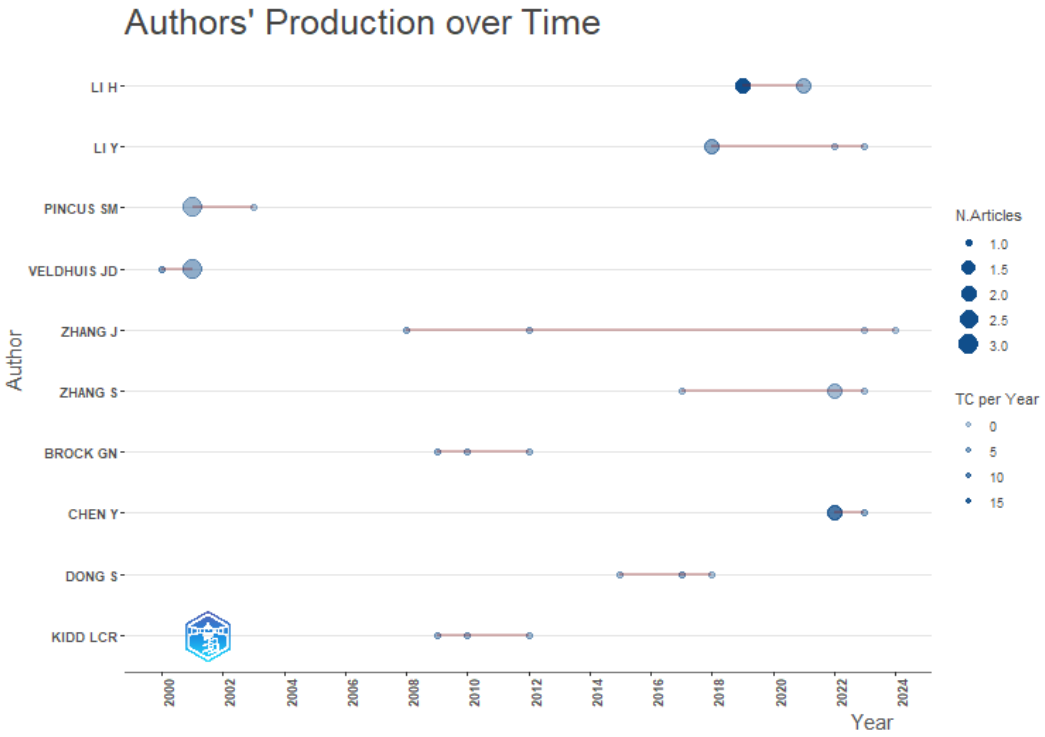
Bu çerçevede yıllar bazında bakıldığında en çok bilimsel üretim 16 tane ile 2023 yılı olurken, ardından 13 adet ile 2022 yılı gelmektedir. 2024 yılları içerisinde ise henüz 6 adet bilimsel üretim bulunmaktadır.

Şekil 2.12'de ise incelediğimiz kısım alanındaki etkin yazarların hangi dergilerde hangi anahtar kelimelerle kullandığının incelenmesini gösterilmektedir.

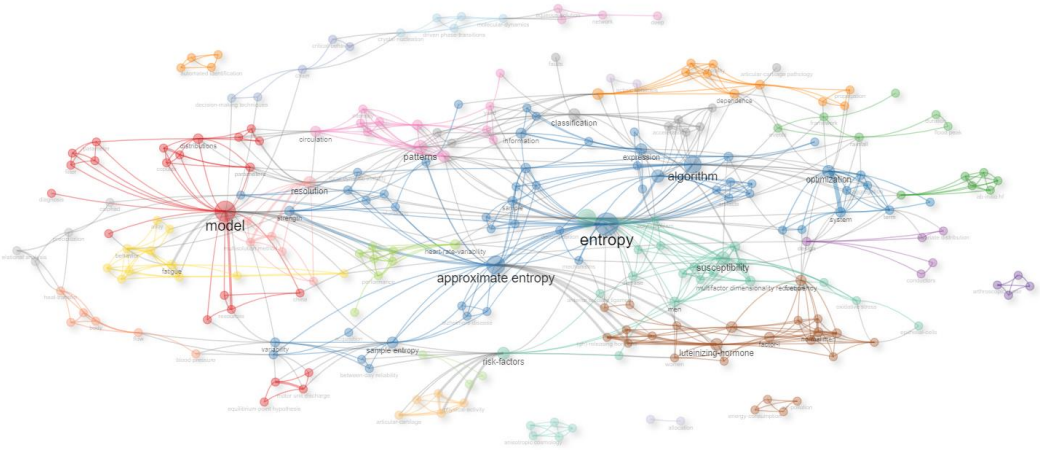


Şekil 2.12 Faktör analizi ve ortak entropi kavramlarına ilişkin dergi, yazar ve anahtar kelime ilişkisi

Şekil 2.12 incelendiğinde ise Faktör analizi ve ortak entropi kavramlarının konusuna ait en ilgili yazarlar listesi bulunmaktadır. Buradaki yazarların zaman içindeki üretimi ise Şekil 2.13'te gösterilmiştir.

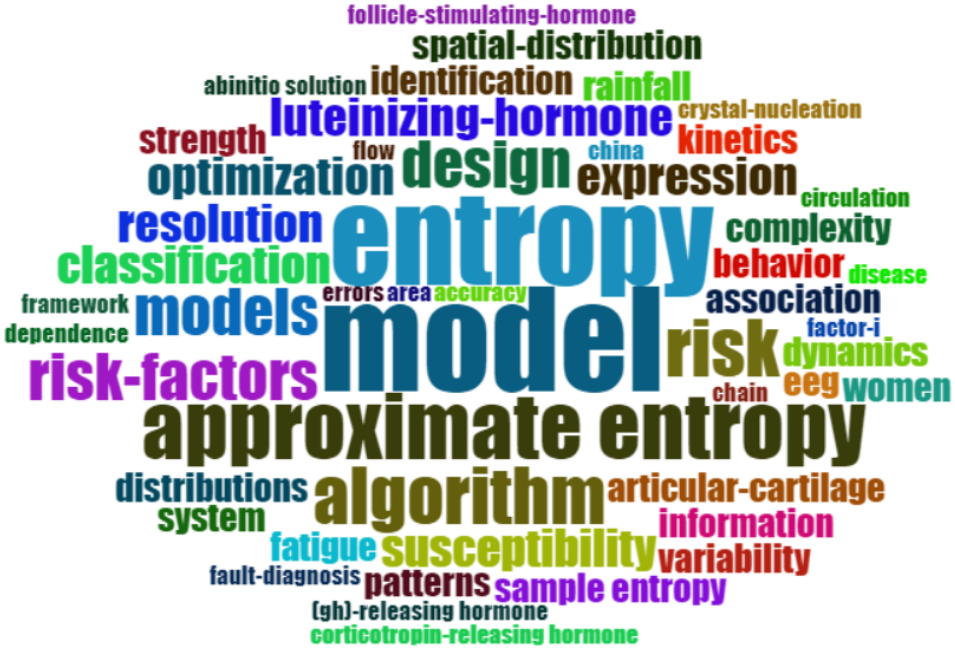


Şekil 2.13 Faktör Analizi ve ortak entropi kavramlarında en ilgili yazarlar



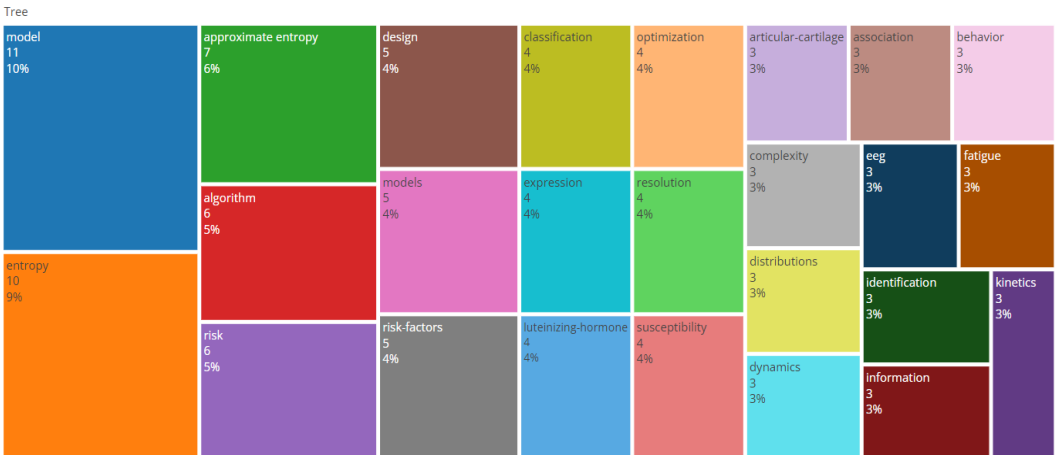
Şekil 2.14 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimeler

Şekil 2.14'te ise "factor analysis" ve "joint entropy" kavramlarının geçtiği yayınlarda, en sık geçen kelimelerin kümelenmesi gözükmektedir. Şekil 2.15'te ise kelime bulutu gözükmektedir. Bu kısımda anlaşılan ise dökümanlarda en sık tekrar eden kelimelerin kelime bulutu ile görselleştirilmiş şeklidir.



Şekil 2.15 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimelere ait kelime bulutu

Şekil 2.16'da ise en sık kullanılan anahtar kelimelerden ilk 15 tanesinin kaç tane bulunduğu gösterilmiştir.



Şekil 2.16 Faktör analizi ve ortak entropi kavramlarında en sık geçen kelimelerin sayısı

2.2 Entropi

Bilim insanları, istatistiksel gözlemler veya deneylerde bilgi arayışından yararlanır. Claude Shannon, bu bağlamda öne çıkan bilim adamlarından biridir. Bu çerçevede bakıldığında, 1948 yılında Claude Shannon'ın "A Mathematical Theory of Communication" adlı kitabında istatistikte entropi kavramı ilk kez ortaya çıktı [2]. Bilgi teorisinde bu kavram Shannon entropisi olarak da bilinir. Shannon'ın makalesi yazılmadan önce, bilgi genellikle belirsiz bir sisli akış olarak görülüyordu. Ancak Shannon'ın makalesi ile, bilginin iyi tanımlanmış ve en önemlisi ölçülebilir bir nicelik olduğunu gösterdi [2].

Boltzmann tarafından ele alınan entropi terimi, 1948 yılına kadar fizik alanında kullanılan bir kavram olmuştur [12]. Bell Laboratuvarlarında bu tarihlerde "Enformasyon Teorisi" üzerinde çalışan Claude Shannon, yüzyılın en ünlü matematikçilerinden biri olan John von Neumann'ın da yardımıyla bu terimi bilgideki belirsizliğin ölçüsünü göstermek için kullanmıştır [12].

2.2.1 Kavramsal Temeller

Farklı birçok bilim insanının araştırmalarıyla entropi kavramı birçok bilim insanı tarafından tanımlanmıştır. Literatürde, termodinamik entropisi ile bilgi entropisi arasındaki ilişki sıklıkla tartışılmaktadır [2]. Bir sistemdeki iş yapamayan enerji miktarı ve bir mesajın taşıdığı bilgi miktarının entropi terimleriyle aynı şekilde ifade edilebilmesi, bu iki ölçüm için kullanılan programlamaların oldukça benzer olmasından kaynaklanmaktadır. Bilgi teorisinde entropi, bir alıcı ile bilgi kaynağı arasındaki iletişim sırasında kaybedilen bilgi miktarını ölçen bir kavram olarak kabul edilmektedir [2]. 1948 yılında Claude Shannon tarafından bir mesajın içerdiği bilgi miktarının ölçülmesi için bu fikir ortaya atılmıştır. Shannon, bir mühendis olarak çalıştığı dönemde, özellikle gürültülü kanallarda sinyal iletimi esnasında bilgi iletimiyle ilgili sorunları incelemiştir. Shannon, gönderilen, alınan ve kaybolan bilgi miktarının ölçülmesine odaklanmış olup bu alanda önemli çalışmaları ortaya koymuştur [2].

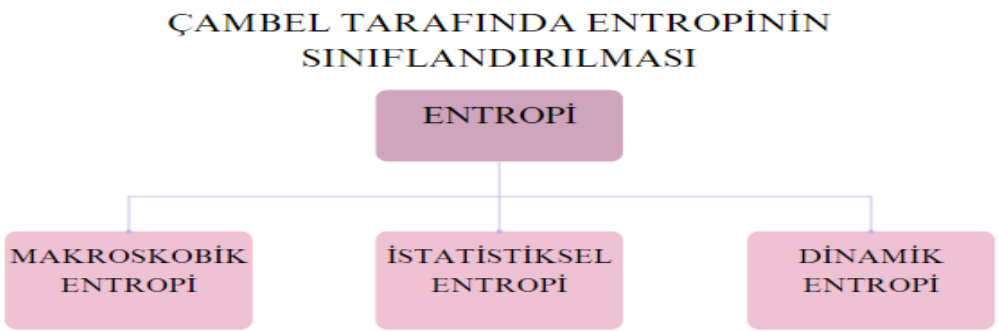
Entropinin tanımı için ortaya atılan her fikir her alanda düzensizliğin ölçülmesi amacıyla kullanılması için tanımlanmıştır [13]. Bilgi teorisi üzerindeki tanımında ise bir veri kaynağındaki belirsizlik (rastgelelik) miktarını temsil eden bit sayısıdır [13].

İletişimin istatistiksel teorisi, Tuller tarafından ortaya atılan bilgi teorisi olarak adlandırılır. Pierce, iletişim teorisi ile bilgi teorisinin birbirleriyle aynı şey olduğunu ifade eder [6]. Gilbert'in aktardığına göre bilgi, mühendisin deney tasarımcısı olarak rolünde kullanışlı olan bir tür zaman veya maliyet ölçüsü olacaktır [6].

İstatistikte entropi kavramını, bilgi teorisinden türetilmiştir. Başlangıçta entropi kavramı termodinamikte düzensizliği veya belirsizlik ölçümü olarak ortaya çıkmıştır. Bilgi teorisi, ele alınacak şekliyle, matematiksel olasılık ve istatistik teorisinin bir alt dalıdır. Bu bağlamda, soyut matematiksel formülasyonları, herhangi bir olasılıksal veya istatistiksel gözlem sistemi için genel geçerli olan bir nitelik taşımaktadır. Dolayısıyla, bilgi teorisi, olasılık ve istatistik gibi çeşitli disiplinlerde uygulanabilmektedir. Modern iletişim teorisinde önemli bir rol oynar, iletişim sistemini stokastik veya rastgele bir süreç olarak formülize eder. Tuller, iletişim teorisinin genellikle bilgi teorisi olarak adlandırıldığını belirtir [6]. Rothstein, bilgi teorisini ölçülebilir kümelerle ilgilenen, belirli bir doğası belirtilmemiş alternatifler arasındaki seçimlerle ilgilenen soyut matematik şeklinde tanımlamıştır [6]. Bilgi teorisinin temel matematiksel ve istatistiksel doğası, gelişimine ve uyarlanmasına büyük ölçüde katkıda bulunan Fisher, Shannon ve Wiener tarafından vurgulanmıştır.

2.2.2 Entropinin Sınıflandırılması

Entropiye dair araştırma yapan birçok bilim insanı farklı tanımlar sunmuştur. Bu tanımları tek bir tabloda düzenlemek için Çambel tarafından sınıflandırma sistemi geliştirilmiştir [14]. Bu tablo ise Şekil 2.17'de gösterilmektedir . Bu bölümde ise bu sınıflandırmayı ve içerdiğini entropi tanımlarıyla ayrıntılı bir şekilde incelemektedir.



Şekil 2.17 Çambel tarafından yapılan entropi sınıflandırması

2.2.2.1 Makroskobik Entropi

Bu sınıflandırmadaki alt başlıklardan olan makroskobik entropi dalında Rudolf Clausius'un entropisi ile Prigogine'nin entropisinin tanımları ele alınacaktır. Prigogine entropisi, Clausius entropisinin genişletilmiş formunu yansıtmaktadır. 1850 yılında Rudolf Clausius termodinamiğin ikinci yasası fikrini öne sürdü ve 1865 yılında ise Rudolf Clausius termodinamiğin ikinci yasası olarak bilinen entropi fonksiyonu kavramsal yaklaşımını ortaya koydu [5].

Prigogine, Clausius'un ileri sürdüğü entropi teriminin açık sistemler için genişletilmiş yeni bir bakış açısını öne sunmuştur. Genel olarak günlük yaşamdan da bilinen bir husus olarak, enerji, madde, bilgi ve entropi çevre ile açık sistemlerle birlikte aralarında sürekli bir etkileşim içinde, sistem içinden dışarıya veya dışarıdan içeriye doğru akar [5]. Bu tür sistemler içinde, entropinin artışı gösterdiği sistemler, Clausius'un tanımlamış olduğu dağınık yapıları sergilemektedir. Causius, yalıtılmış bir sistemdeki entropi artışının nasıl hesaplanacağını açıklamıştır.

Ek olarak Prigogine, sistem ile çevresi arasında entropi değişimini de odak alarak daha derinlemesine bir inceleme yapmıştır.

2.2.2.2 İstatistiksel Entropi

Bu başlık Boltzmann, Schrödinger ve Shannon entropisinin tanımlarını kapsamaktadır.

Boltzmann'ın dinamik ve termodinamik alanlarındaki bağlantıyı bulmak bağlamında başlattığı çalışmalar, 1948'de Shannon'un Enformasyon Teorisi ile ilgili olarak elde ettiği olasılık temelli sonuçlarla belirgin bir benzerlik sergilemektedir. Bu durumda her iki yöntem de istatistiksel bir karakteristik sergilemektedir. 1933'te Nobel Fizik Ödülü kazanmış olan Erwin Schrödinger, entropi hakkında başka bir entropi yorumu da mevcuttur.

Schrödinger, canlı organizmaların negatif entropi veya bir diğer ifadeyle negentropi ile beslendiklerini iddia etmiştir. Bir canlı organizmanın termodinamik denge durumundaki bozunmasını (ölüm halini) geciktiren olağanüstü yeteneğini, istatistiksel teori ile açıklamaya çalışan Schrödinger'in aktardığına göre, canlı bir organizma, yaşamak ve kendisini sabit ve göreceli olarak düşük bir entropi düzeyinde tutabilmek için negatif entropiden yararlanmaktadır [5].

Shannon entropisi, ileriki kısımlarda daha ayrıntılı bir biçimde ele alınacaktır. Bu konuda daha derinlemesine bir inceleme gerçekleştirilecektir.

2.2.2.3 Dinamik Entropi

Şimdiye kadar yapılan sınıflandırmalarda, entropi tipik olarak rasgeleliğin belirleyicisi olarak kabul edilmiştir. Bununla birlikte, bu bölümde, zamana bağlılık ve dinamik karakteristikler üzerine çalışmalar yapan iki önemli bilim insanı ele alınmıştır.

Andrey Kolmogorov, metrik-entropi olarak da bilinen teorik-ölçüm entropisi literatürde entropi tanımının kazandırılmasında önemli bir rol oynamıştır [5]. Dinamik entropi, sıralı olasılıkların kullanıldığı, oransal olarak belirtilen ve ek olarak limit durumlarıyla aracılığıyla ölçülen entropinin tanımını içermektedir [5]. Shannon tarafından ortaya atılan entropinin bir parametreye bağlı olarak genelleştirilmiş hali Alfred Reyni tarafından yapılmıştır ve bu sayede Renyi'nin çalışmaları, entropiyi daha geniş bir çerçevede ele almayı olanaklı bir hale getirmiştir.

Alfred Renyi, 1961 yılında biyolojik çeşitliliğin ölçümünde kullanılabilecek bir kavram üzerinde çalışmalar yapmmıştır ve renyi entropisini adı altında ileri sürmüştür [5]. Reyni entropisi kuantum bilgi ölçümü alanında da oldukça önemli bir rol oynamasının yanı sıra Shannon entropisinin özel bir durumudur.

2.2.3 İstatistiksel Belirsizliğin Ölçülmesi

Modern determinizm anlayışına göre, istatistik alanını matematiğin uygulamalı bir dalı olarak kabul eder. Fizikten biyolojiye birçok bilim alanlarında sıklıkla kullanılır. Temelde bu olasılık teorisinden doğan ve matematiksel modeller için tümdengelime bir yaklaşımın kullanıldığı bir yöntemdir [3]. Matematiksel modellerin gelişimiyle başlayan bu süreçte tümdengelim bir yaklaşım ile belirli varsayımlardaki uygunluk şartları altında gözlenen değerlerin ortaya çıkma olasılıklarının tersine bir çıkarsama yöntemi kullanarak hesaplanır. Ek olarak gerçeği tam anlamıyla algılanmasını kısıtlayan unsurlar sebebiyle faktörlerin olumsuz etkilerinin azaltılmasına yönelik bir hedef içerisindedir.

N sayıda gözlem ve n sayıda değişkenin olduğu bir düşünüldüğünde, her değişkeni N boyutlu bir vektör şeklinde düşünebiliriz ve böylece onları bir uzayda yer alan noktalar olarak algılamamızı sağlayabilir[3]. Bu bağlamda belirtilen sebepler

doğrultusunda, sözü edilen vektörlerin uzaydaki konumları kesin olmamakla beraber olasılıklı yapıya sahip olacaktır. Bu bağlamda vektörlerin uzaydaki yerleri, iki durum şeklinde inceleme halinde ele alınabilir.

İlk durum için incelediğimizde, tüm değişkenlerin arasındaki belirsizlik durumunun maksimum olduğu halidir. Bu durumda vektörlerin bütünü tümüyle rasgele şekilde ve düzensiz olarak dağılmıştır. Bu durum ise değişkenlerin arasındaki belirsizlik seviyesinin maksimum olduğunu ifade eder [3]. İkinci durumu incelediğimizde ise, vektörlerin bütünü belirli yerlerde yoğunlaşmıştır. Bu durum ise belirli bir düzen içerisinde olduğunu ifade etmektedir. İki durumun karşılaştırılması yapıldığında ilk durumda değişkenler arasındaki belirsizlik seviyesi en üst seviyedeyken, ikinci durumda ise değişkenler arasındaki belirsizlik en düşük seviyededir [3]. Ancak gerçek durumlar bu iki uç durumun arasında yer almaktadır.

İstatistik alanı fizik alanı ile benzerlik kurması amacıyla değişkenler, fiziksel bir sistemin makroskopik değişkenlerinin ele alınmasıyla istatistikteki belirsizliğin entropi ile aynı olduğu görülmektedir. İstatistiksel bir kavram olan parsimoni terimi entropi ile bağdaştırılan bir istatistiksel belirsizlik tanımına karşılık gelmektedir [3]. Bu bağlamda bu iki kavramın arasındaki benzerlik oranı oldukça yüksektir. Parsimoni kavramı genel anlamda en az aksiyom, varsayım ile bir teoremin kurulmasını ifade etmekle beraber istatistikte parsimoni ise değişkenlerin araasındaki ilişkinin en az sayıda denklemle veya en az sayıdaki açıklayıcı değişken ile ifade edebilmesini ifade etmektedir [3].

2.2.4 Bilgi Teorisinde Entropi

Bir sistemin düzensizliğinin ölçülmesinde bilgi teorisindeki entropi kullanılmaktadır. Bilgi teorisinde entropi kavramı H ile ifade edilmektedir [13]. Entropinin kullanım alanı olarak bir bilgi ortamındaki düzensizlik seviyesini belirlemek amacıyla kullanılmaktadır. Bilginin monoton bir düzen içerisinde akması durumunda düzensizliği azaltabilir. Entropi formülündeki n genellikle veri miktarını temsil etmektedir [13]. Ek olarak p_i ise ilgili verideki olasılığı temsil etmektedir.

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{2.1}$$

(2.1) ile ifade edilen formül esas alındığında, p ve 1-p olasılık değerlerine sahip olan iki durum varsayımı altında, bu iki olasılığın eşit olduğu durumda entropi değerinin maksimum olduğu noktadır [13].

Shannon teorisine göre, bir sistem içinden alınacak kararların güvenilirliğinin ve doğruluğunun belirlenmesindeki kritik unsur, o sistem ile ilgili sahip olunan bilgi miktarı ve kalitesidir.

Bu çalışmada asıl ilgilenilen konu bilgi teorisi, olasılık, istatistik ve matematik teorilerinin bir dalıdır. Bu sebeple bilgi teorisini bu olanlarda oldukça sık görmekteyiz. Ek olarak modern iletişim teorisinde önemli bir rol oynamakla birlikte iletişim sistemini stokastik veya rastgele bir yöntem olarak formüle eder.

Bir sistemde girilen bilgi miktarındaki artış, sistemdeki belirsizlik seviyesinin azalması sebebiyle entropi değerinin düşmesine sebep olacaktır. Bunun tersi olarak diğer taraftan ise bilgi miktarındaki azalış, belirsizlik seviyesinin artması sebebiyle entropi değerinin yükselmesine neden olacaktır. [13]

Bir sistemin bilinmeyen yapısındaki artan bilgi miktarı, sistemin belirsizliğini azaltacaktır. Bu bağlamda istatistiksel bir bakış açısından ele alınan kutulardan top çekme örneğinde ve zar atma deneyinde bilgi miktarları ile entropi düzeyleri arasındaki ilişki, örneklerde detaylı bir şekilde incelenmiştir [15].

Örneğin olasılık teorisine dayalı olarak, kutudan top çekme deneyini ele alalım. Varsayalım ki bir kutuda 4 top bulunmaktadır ve bu deney, kutudan çekilen topun rengine bağlı olarak, kırmızı veya mavi top gelme olasılığını incelemektedir. Öncelikle, kutuda bulunan 4 topun tamamının kırmızı olduğunu düşünelim. Bu durumda, kırmızı bir top çekme olasılığımız P(K) = 4/4 olarak belirlenir ki bu durumda kutudan çekilen topun kırmızı olacağı kesindir. Bu durum, yüksek bir bilgi miktarına sahip olduğumuzdan entropi düzeyinin düşük olduğu bir durumu ifade eder.

Başka bir senaryo olarak, kutuda 2 kırmızı ve 2 mavi top olduğunu düşünelim. Bu durumda, kutudan çekilen bir topun renginin kırmızı ya da beyaz olma olasılıkları eşit olacaktır, yani P(K) = 1/2 ve P(M) = 1/2. Bu durumda, kutudan çekilecek olan bir topun rengi hakkında düşük bir bilgi miktarına sahibiz. Bilgi düzeyinin düşük olması nedeniyle entropi seviyesi yüksektir.

Deneyi 3 kırmızı ve 1 mavi top içeren bir durumu ele alarak sürdürelim. Bu durumda, kırmızı top gelme olasılığı P(K) = 3/4 ve mavi top gelme olasılığı P(M) = 1/4 olarak belirlenir. Kutudan bir top çekildiğinde kırmızı olma olasılığının daha yüksek olduğu bir durumla karşı karşıyayız. Bu durum, kutudan çekilen topun renginin kırmızı olma olasılığının daha yüksek olduğu anlamına gelir. Dolayısıyla, kutudan çekilen topun rengi konusundaki bilgi miktarının orta düzeyde olduğunu söyleyebiliriz. Bu durum, orta düzeyde bir bilgi miktarına sahip olduğumuzdan entropi seviyesinin orta düzeyde olduğu bir durumu yansıtır.

Başka bir örnek ele alınması gerekirse istatistiksel bir çerçevede ele alınan zar atma deneyinde, bir zarın atılmasıyla elde edilebilecek altı sonuç bulunmaktadır. Normal şartlar altında bir zar atıldığında, herhangi bir yüzün gelme olasılığı 1/6'dir. Ancak, bir zar atıldığında çift sayı gelme olasılığını, yani istenilen durum 1/2 olduğu açıktır.

Bu bilgilerin yanı sıra, hileli bir zarın kullanıldığı bir durumu ele alalım. Varsayalım ki hileli bir zar kullanıldığında, üst yüzeye gelen sayının çift sayı olma olasılığı, hilesiz zara kıyasla daha yüksek olsun. Bu durumda, hileli zarın daha yüksek olasılığa sahip olduğu gözlemlenmektedir. Bilgi miktarlarına dayalı olarak bu iki durumun entropilerini karşılaştırdığımızda, hileli zardaki daha yüksek olasılık nedeniyle entropisi, hilesiz zara göre daha düşüktür. Sonuç olarak, hilesiz bir zar atma deneyinde entropi daha yüksek olacaktır.

Diğer bir örnek ele aldığımızda ise düzgün bir paranın üç kez atılışındaki olasılıklara sahip, 8 durumlu bir sistemin entropisini hesaplansın. Paranın üç kez atılması durumunda, 8 farklı sonuç olasıdır çünkü her atışta iki olası sonuç vardır (tura veya yazı). Bu durumda, 2 üstü 3 yani 8 farklı kombinasyon elde edilir. Bu durumu daha iyi anlamak için her bir atışın olası sonuçlarını düşünelim:

- Tura - Tura - Tura
- Tura - Tura - Yazı
- Tura - Yazı - Tura
- Tura - Yazı - Yazı
- Yazı - Tura - Tura
- Yazı - Tura - Yazı
- Yazı - Yazı - Tura

- Yazı - Yazı – Yazı

Bu 8 durumlu sistemin entropisini hesaplamak için (1)deki formülü kullanabiliriz:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.2}$$

Formülde H(X) entropiyi pi ise her bir durumun olasılığını ifade eder. Bu durumda, her durumun olasılığı eşit olduğu için her bir $p_i = 1/8$ olacaktır.

Şimdi, sistemin entropisini belirlenebilir.

$$H = -8\left(\frac{1}{8} log_2 \frac{1}{8}\right) = 3(bit) \tag{2.3}$$

2.2.5 Entropinin Özellikleri

Entropi terimi bilgi teorisinde tanımlanmasında belli başlı özellikler altında şekillendirilmiştir [13].

Bu bağlamda özelliklerin ilkinde eğer bir sistemin entropisi sıfırsa, istisna bir durum haricinde tüm durumların olasılıklarının sıfır olduğu belirtilebilir. Bu istisna durum için ise tek bir durum kalmaktadır. Bu durum ise, tüm olasılıkların tamamına sahiptir.

Diğer bir özellik ise her bir p_i değeri belirli n için eşit olduğu zaman entropi en yüksek değere ulaşır ve log(n)'e eşittir. Bu özellik sezgisel açıdan en belirsiz olan durum olarak ifade edilmektedir.

Üçüncü özellikte ise ele alınacak bir diğer nokta ise x ve y olacak şekilde iki olay ele alındığında ve bu olayların sırasıyla olasılık değerleri m ve n olarak beliritildiğinde;

Ortak olasılığı $p(\mathrm{i}, j)$'dir. Bu ortak oluşumun entropisi;

$$H(x,y) = -\sum_{i,j} P(i,j)\, log_p(i,j) \tag{2.4}$$

$$H(x) = -\sum_{i,j} P(i,j)\, log_p(i,j) \tag{2.5}$$

$$H(y) = -\sum_{i,j} P(i,j)\, log_p(i,j) \tag{2.6}$$

Bağımsız olaylar söz konusu olduğunda, bireysel entropilerin toplamının ortak olayın belirsizliğine eşit ya da daha büyüktür şeklinde ifade edilebilir.

$$H(x,y) \leq H(x) + H(y) \tag{2.7}$$

$$p(\mathrm{i},j) = p(\mathrm{i})p(j) \tag{2.8}$$

Öne çıkan dördüncü özellik olarak {1,0,0,…,0}, {0,1,0,…,0},…{0,0,0,…1} şeklindeki dağılımların entropi değeri 0 dır.

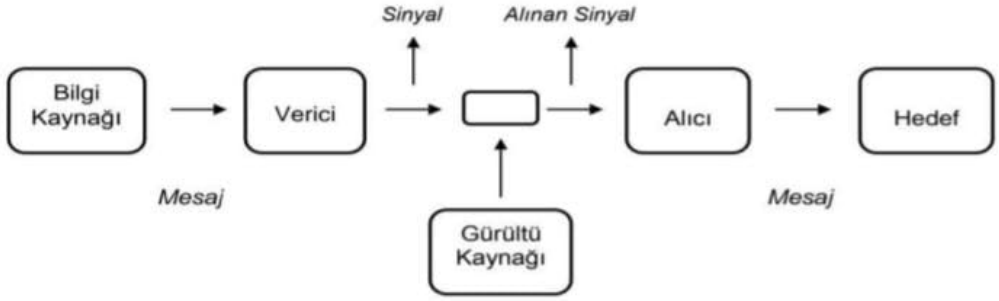
Bir başka özelliğe değinmek gerekirse bir olasılık dağılımındaki tüm pi değerlerinin birbirine eşit olmasına yönelik bir değişim entropi değerini arttırır.

2.2.6 Shannon Entropisi

Claude Elwood Shannon, 30 Nisan 1916 yılında Amerika'nın Michigan eyaletindeki Gaylord kasabasında, Claude Elwood ile Mabel Wolfnon'un oğlu olarak doğmuştur. Shannon, elektronik iletişim çağının babası olarak tanınmaktadır. Bunun sebebi ise bilimsel anlamda bilgiyi tanımlamış ve bilgi ölçümünün mümkün olduğunu gösteren bir çalışmaya imza atmıştır. Bu kavramı "A Mathematical Theory of Communication" başlıklı araştırma makalesi sayesinde öne sürmüştür. Genç yaşlarında Gatlord'da yaşayan Shannon 1932 yılında Gaylond Lisesinden mezun olmasının ardından Michigan Üniversitesine katılmıştır. 1936 yılında matematik ve elektrik mühendisliği alanlarından derece almıştır. Shannon Massachusetts Institute of Technology (MIT)'ye katılarak elektrik mühendisliği alanında yüksek lisans derecesine sahip olmuştur. Yüksek lisans tezi "A Symbolic Analysis of Relay and Switching Circuits" başlığını kullanmıştır. Tezinde ise boole cebiri kullanarak röle anahtar devrelerini analiz edilmesi ve optimize etmesi konusunu ele almıştır. Yüksek lisans eğitimini elektrik mühendisliği alanında tamamlayan Shannon doktora eğitimini 1940 yılında matematik alanında bitirmiştir. Doktora tezinde ise popülasyon genetiği üzerinde durmuştur. 1942 yılında Bell Telefon Laboratuvarları'nda çalışmaya başlamıştır. Shannon buradaki çalışmaları sırasında bilgi iletişiminin açıklamasını içeren bir teori formüle etmesinin yanı sıra bilginin en etkin biçimde iletilmesi için üzerinde çalışmalar yapmıştır. [16]

Claude Elwood Shannon, 24 Şubat 2001 yılında ölmüştür. Kendisi bir matematikçi, elektrik mühendisi, bilgisayar bilimcisidir ve "bilgi teorisinin babası" olarak isimlendirilmiştir. 1948 yılında yayınlanmış olan "İletişimin Matematiksel Teorisi" isimli makalesinde Shannon, "bilgi entropisi" olarak da bilinen ve aynı zamanda bir diğerr adı Shannon entropisi olan kavramı ortaya atmıştır. Enformasyon kuramı çerçevesinde yeniden bir değerlendirme yapan Shannon, ortalama bilgi adı altında istatistiksel entropiyi tanımlamış olan kişidir. Bu yaklaşım sayesinde modern iletişim teorisinin temellerini atmıştır. [16]

Shannon makalesinde bilgi kuramı alanındaki çalışmalarında ortaya atmış olduğu iletişim sistemi Şekil (2.1) ile gösterilmiştir.



Şekil 2.18 Shannon'nun tanımladığı iletişim sistemi

Şekil 2.18'de gösterilmiş olan Shannon'nun tanımladığı iletişim sisteminde bilgi kaynağı, alıcı birime iletilmesi amacıyla ileti ya da ileti dizisi üreten sistemdir. Verici ise, bilgi kaynağı ile üretilmiş olan mesajı, kanal boyunca iletim için uygun sinyal oluşturan bir sistemdir. Mesajı kanalda iletmek amacıyla kullanılan biçime sinyal adı verilmektedir. Vericiden alıcıya iletilen sinyalin ortamı kanal olarak adlandırılmaktadır. Alıcı ise bu sinyali yeniden mesaj formatına dönüştürür ve genellikle vericinin yaptığı işlemin tersini yapmaktadır. Alıcı terminal ile iletişim kurmak için bir ileti ya da ileti dizisi üreten sistem, bilgi kaynağı olarak isimlendirilmektedir. Sinyalin vericiden alıcıya iletilmesinde kullanılan ortam kanal olarak adlandırılmaktadır. Bilginin varış yeri yani bir diğer adı hedef, iletinin ulaştığı son noktadır şeklinde tanımlanabilmektedir. [17]

İletişim sistemleri ilgilendiren belirli genel sorunları ele almak gerekirse öncelikle sinyalin kanalda iletimi esnasında ortamın özellikleri sebebiyle birtakım değişiklikler meydana gelebilmektedir. İletinin iletim süreci boyunca sinyalin

geçtiği ortamın özellikleri nedeniyle istenmeyen bu değişimlere gürültü adı verilmektedir. Eğer ki bir iletişim sürecinde bir gürültü mevcutsa, kaynaktan gelen ileti ile bilginin varış yerine gelen ileti arasında farklılıklar olabilir. Bu sorun en sık görülen problemlerden biridir. Öte yandan, bir iletinin bir alıcıya herhangi bir değişiklik olmaksızın ulaşım ihtimali oldukça düşüktür. İletilen bir iletinin zamanla bozulma meydana gelmesi durumunun bir türünü Shannon, entropi olarak tanımlamıştır. Bununla birlikte, gürültünün mevcut olduğu bir iletim kanalı aracılığıyla, çok az bir hata ile bilginin iletilmesinin mümkün olabileceğini göstermiştir. Böylece Shannon'nun bilgi kuramı sayesinde bilginin en doğru bir şekilde ve en hızlı olacak biçimde iletilmesine yardımcı olabilmektedir. [2]

Shannon'nun entropisi iletişim kuramı açısından bir ölçü olarak tanımlanmış olmasına rağmen, bu ölçünün geniş çaplı bir kullanım alanı mevcuttur. Bu alanlardan bazıları matematik, mühendislik, fizik, biyoloji, sosyal bilimlerdir. Entropinin bu önemli başarısı, özellikler olasılıksal sistemde belirsizliği ölçebilmesindeki olarak söylenebilmektedir.

Shannon, bir olay hakkındaki bilgi edinme sürecini, olayın belirsizlik içermesi gerektiği durumlarda gerçekleşeceğini belirtmiştir. Bu durumlarda, yüksek olasılıklı olayların ortaya çıkması genellikle çok az bilgi verir, ancak düşük olasılıklı olayların ortaya çıkması genellikle daha fazla bilgi verir. Bu bağlamda, belirli bir alternatifin gerçekleşme olasılığı, ilgili çözümü gösteren işaret, sembol ya da sayının belirsizliğini göstermektedir. Sonuç olarak, edinilen bilgi, giderilen belirsizliğin miktarını dolaylı olarak ölçer. [18]

Entropi terimini bu yöntem ile Shannon, bir olayın çeşitli seçenek durumlarının beklenen değeri olarak matematiksel bir ilişkiyle tanımladı. Entropi bit veya desibel gibi birimlerle ölçülebilen bir nicel büyüklüktür. Bu sebeple bir rasgele sürecin olasılıklı yapısının belirlenmesinin ardından bu sürecin entropisi hesaplanır ve belirli birimler ile ifade edilebilmektedir. Shannon geliştirmiş olduğu entropi formülündeki logaritmayı 2 tabanında kullanmıştır. Logaritma kullanılmasının amacı ise bir bağımsız sistem içindeki entropinin toplanabilir nicelik formatında getirilebilmesi için kullanılmaktadır. Bu kapsamda, logaritma işleminin bilimsel çalışmalarda da sayısal kodlama sistemlerinde de uygun olduğu kanısına varılmıştır ve bu sebeple iki tabanında logaritma kullanılmıştır. İki tabanında logaritmanın tercih edilmesinde en büyük etken ise sayısal kodlama sistemlerine uygun

olmasından kaynaklanmaktadır. Ek olarak entropi değerlerinin hesaplanmalarında elde edilen sonuçlarının ikili sayı sisteminde kısaltılarak ifade edilmesine olanak sağladı. Bu şekildeki kısaltma birimine bit denmektedir. [18]

Olası durumları $p_1, p_2, p_3, \dots, p_n$ kümesi olarak tanımlayan Shannon, bu kümedeki hangi olayın gerçekleşeceği bilinmediği için bu oluşan belirsizliği entropi yardımıyla ölçülebileceğini ifade etmiştir. Entropinin genel formülü (2.8) eşitliğinde gösterildiği gibidir. Denklemdeki H entropiyi ifade ederken, k sabiti boltzman sabiti olarak tanımlanmaktadır. [6]

$$H = -K\sum p_i \log_2 p_i \tag{2.8}$$

Shannon, k sabitini 1 kabul ederek ortaya Shannon entropisi ya da bir diğer ifadeyle Shannon belirsizliğini ortaya koymuştur. Shannon entropisi (2.9) denklemi ile ifade edilmektedir.

$$H = -\sum p_i \log_2 p_i \tag{2.9}$$

Bu bağlamda bu ifadeler ışığında varılan sonuç, rasgele bir değişkenin değerini tahmin ederken ortaya çıkmış olan belirsizliğin sayısallaştıran entropi değeridir. Özetle Shannon belirsizliği bilgi teorisinde kullanılmaktadır ve olasılık dağılımının sahip olduğu belirsizliği ölçer.

Bu bağlamda örnek vermek gerekirse istatistik alanında sıklıkla kullanılan para atma deneyi ve zar atma deneyi için entropi değerlerinin karşılaştırılmasına dair bir örnek kullanılsın [15]. İlk olarak bu iki deney içinde örneklem uzayının belirlenmesi gerekir. Örneklem uzayı bir deneydeki tüm olası durumların temsilidir. Öyleyse para atma deneyinin örneklem uzayı S_1= {Y, T} ve zar atma deneyi için örneklem uzayı S_2 = {1, 2, 3, 4, 5, 6} şeklindedir.

Bu iki deney için örneklem uzayına bakıldığında ve iki ayrı sistem olarak düşünürsek olası sonucun daha fazla olduğu zar atma deneyinde daha fazladır. Böylelikle bu durumdaki belirsizliğin diğer deneye kıyasla daha fazla olduğu kanısına varırız. Belirsizliğin daha fazla olması halinde entropi değerinin daha yüksek olduğunu ifade edilebilir.

Bu gözlemi Shannon belirsiziliği ile doğrulanması için (2.10) denkleminden faydalanrak bulabiliriz.

$$H(x) = -\sum_{i=1}^{n} p_i \, ln \, p_i \qquad (2.10)$$

İlk olarak para atma deneyinin örneklem uzayı S_1= {Y, T} şeklindeydi. Bu durumda n=2 denilir. Olasılık değerleri ise P(Y) = P(T) = 1/2 yani p_1 = p_2 = 1/2 olarak gösterilir.

O halde söz edilen veriler dikkate alındığında para atma deneyinin entropi fonksiyonu (2.10) formülünden faydalanarak aşağıdaki gibi sınuca ulaşırız:

$$H(x_1) = -\sum_{i=1}^{2} p_i \, ln \, p_i = -(p_1 \ln p_1 + p_2 \ln p_2) = -\left(\frac{1}{2} ln \frac{1}{2} + \frac{1}{2} ln \frac{1}{2}\right)$$

$$= -[0.5(-0.69) + 0.5(-0.69)] = 0.69$$

Öyleyse H(x_1) = 0.69 olduğu açıkca gözükmektedir. O halde para atma deneyi için sonucundaki belirsizliğin ölçüsü, Shannon entropi fonksiyonu ile hesaplanan değer 0,69 değerini vermektedir.

Zar atma deneyine baktığımızda ise S_2 = {1, 2, 3, 4, 5, 6} şeklindeydi. Bu örneklem uzayına bakıldığında n=6 olduğu gözükmektedir. Olasılık değerleri ise p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6 olarak ifade edilir.

O halde söz edilen veriler dikkate alındığında para atma deneyinin entropi fonksiyonu (2.10) formülünden faydalanarak aşağıdaki gibi sonuca ulaşırız:

$$H(x_2) = -\sum_{i=1}^{6} p_i \, ln \, p_i$$

$$= -(p_1 \ln p_1 + p_2 \ln p_2 + p_3 \ln p_3 + p_4 \ln p_4 + p_5 \ln p_5 + p_6 \ln p_6)$$

$$= -\left(\frac{1}{6} ln \frac{1}{6} + \frac{1}{6} ln \frac{1}{6} + \frac{1}{6} ln \frac{1}{6} + \frac{1}{6} ln \frac{1}{6} + \frac{1}{6} ln \frac{1}{6} + \frac{1}{6} ln \frac{1}{6}\right)$$

$$= 1.79$$

Öyleyse H(x_2) = 1.79 olduğu açıkça gözükmektedir. O halde zar atma deneyi için sonucundaki belirsizliğin ölçüsü, Shannon entropi fonksiyonu ile hesaplanan değer 1.79 değerini vermektedir.

Sonuç olarak para atma deneyi ile zar atma deneyinin karşılaştırıldığında, zar atma deneyinde entropinin değerinin daha yüksek olduğu sonucuna sayısal verilerle ulaşılmıştır. Buradan anlaşılan konu zar atma deneyinde daha fazla olması sonucun olması dolayısıyla belirsizliğin daha fazla olması durumuyla ilişkilidir. Olası sonucu yüksek olan zar atma deneyindeki belirsizlik daha çok olduğu için entropi de daha yüksek çıkmıştır.

Tablo 2.1: Zar atma ile Para atma deneylerinin kıyaslanması

Deney	**Para atma deneyi**	**Zar atma deneyi**
Olası sonuçlar	Yazı, Tura	1, 2, 3, 4, 5, 6
Entropi değerleri H(Xi)	0.69	1.79

2.2.7 Ortak Entropi

Bilgi teorisinin temellerini ortaya atmış olan Claude Shannon, ortak entropi ile Shannon entropisinin arasındaki ilişkiyi açıklamıştır. Bir önceki bölümlerde yazıldığı gibi bir olayın meydana gelme olasılığına göre bulunan bilgi ölçüsü Shannon entropisi ile elde edilmektedir. Bununla birlikte, eğer değişken sayısı iki ya da daha fazla olması halinde rasgele değişkenlerin kendi aralarındaki belirsizliğin ölçümüne ortak entropi denir. Shannon bu kavramlar üzerinde çalışmaları sayesinde gelişen bu alan, günümüzde istatistik, matematik, bilgisayar bilimleri ve birçok başka alanda uygulamalarda kullanılabilmektedir.

Ortak entropinin genel anlamda tanımı $p(x, y)$ olasılık dağılımına sahip ve (X,Y) kesikli rasgele değişkeninin entropisi (2.11) denklemi ile tanımlanır [19].

$$H(X,Y) = -\sum_{x \epsilon X} \sum_{y \in Y} P(x,y)\, log_P(x,y) \tag{2.11}$$

Ortak entropi formülünün daha detaylı görünmesi için sayısal bir örnek ile daha iyi anlaşılması ve buna ek olarak Shannon entropisi ve ortak entropi arasındaki bağlantının açıklanması açısından bir örnek incelemesi yapılabilir. Bu amaç

doğrultusunda ilk olarak iki bağımsız rastgele olayı düşünelim. bunlar A ve B olsun. Her bir olayın gerçekleşme olasılıklarını bilinsin ve bunlar;

P(A = 0) = 0.4, P(A = 1) = 0.6, P(B = 0) = 0.7, P(B = 1) = 0.3 şeklinde olsun.

Öyleyse Shannon entropi değerlerinin hesaplanması için (3) denklemi ile bulunur.

$$H(A)= -(0.4\cdot\log2(0.4)+0.6\cdot\log2(0.6)) = 0.971$$

$$H(B)= -(0.7\cdot\log2(0.7)+0.3\cdot\log2(0.3)) = 0.881$$

Ardından, A ve B'nin birlikte gerçekleşme durumunu ele aldığımız durumun incelensin. Ortak entropiyi hesaplamak için (2.12) formülünden faydalanarak bulabiliriz.

$$\sum_{i,j} p(a_i, b_i)\, log\, p(a_i, b_i) \qquad (2.12)$$

Buradaki formülde A ve B'nin olası değerleri a_i ve b_i oalrak temsil edilir. Bu iki olayın birlikte gerçekleşme olasılığı ise $p(a_i, b_i)$ ile ifade edilir. Bu formülü uygulayabilmek adına ilk olarak A ve B'nin birlikte gerçekleşme durumlarının sırasıyla şu şekilde verildiğini varsayalım:

H(A=0, B=0)= (0.4)(0.7) = 0.28

H(A=0, B=1)= (0.4)(0.3) = 0.12

H(A=1, B=0)= (0.6)(0.7) = 0.42

H(A=1, B=1)= (0.6)(0.3) = 0.18

Son aşamada ortak entropiyi hesaplanması (2.12) formülü yardımıyla aşağıdaki gibidir:

$$H(A,B)= -(0.28\cdot\log2(0.28)+ 0.12\cdot\log2(0.12)+ 0.42\cdot\log2(0.42)+ 0.18\cdot\log2(0.18)) = 1.973$$

Bu örnek, Shannon entropisi ve ortak entropi arasındaki ilişkiyi anlamak için temel bir gösterim sunmaktadır.

2.2.8 Ortak Bilgi

Rasgele değişkenler arasındaki ilişkinin ölçümünde kullanılan yapının adı ortak bilgi olarak isimlendirilmiştir. Bu bağlamda bilgi teorisi kavramının temel kavramlarından biridir. Özellikle iletişim sistemindeki kanal, giriş ve çıkışlarda

oluşan belirsizliğin ölçülmesinde ve bu kanal kapasitelerindeki anlaşılabilirliğin ölçülmesinde fayda sağlamaktadır. Bu bilgilere ek olarak sesin kodlanmasında ve tanınmasında, bunların yanı sıra EEG, MEG gibi biyomedikal sinyallerin işlenmesi gibi çeşitli sinyal işleme alanlarında yaygın olarak kullanılmaktadır. Makine öğrenimi araçları kısmında da önemli bir rol oynamaktadır. [20]

Ortak bilgi metodunun araştırmacılar tarafından tercihinin bir nedeni ise, sadece doğrusal kovaryansa bağlı bağımlılıklarından değil, aynı zamanda doğrusal olmayan ilişkileri de açığa çıkarmasından kaynaklanmaktadır. Bu bağlamda bu metod ile yapılan bir analizde kapsamın genişletilmesiyle, hem doğrusal hem doğrusal olmayan ilişkilerin açığa çıkmasını mümkün kılar. [20]

Ortak bilgi miktarı, entropi ile ifade edilişi ise I(X;Y) = H(X)-H(X/Y) şeklindedir. [21] Bir rastgele değişkenin kendiyle olan karşılıklı bilgisi, rastgele değişkenin entropisi olarak da adlandırılır. Bu durum, entropinin zaman zaman kendi bilgisi olarak da anılmasının temelidir. Bu bağlamda formüllerin tek bir başlık altında toplamak ve karşılıklı bilgi ile entropi arasındaki ilişkiyi görmek gerekirse, bu Şekil 2.19'da gösterilmektedir. [21]

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
I(X;Y) &= H(Y) - H(Y|X) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
I(X;Y) &= I(Y;X) \\
I(X;X) &= H(X).
\end{aligned}
$$

Şekil 2.19 Karşılıklı bilgi ve entropi arasındaki ilişki

2.3 Faktör Analizi

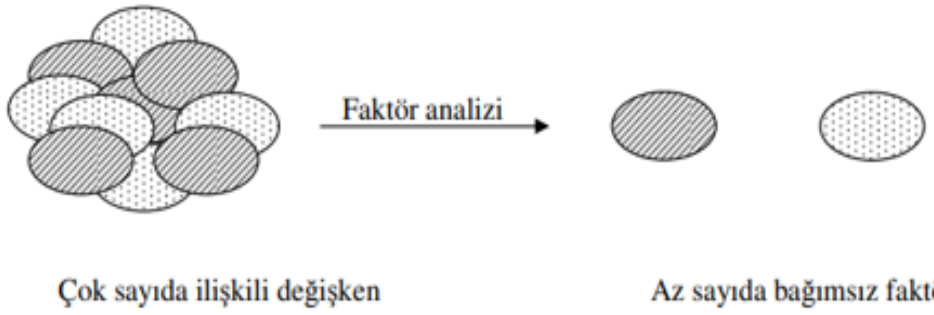
Günümüzde, büyük veri setlerinin kullanılmaya başlanması ve daha derinlemesine analizler yapma ihtiyacının doğması sebebiyle karmaşık yapılı veri setleri ortaya çıkmaktadır. Bu veri setleri ile etkili bir analiz sunmak amacıyla çok sayıda araştırmacı verinin daha anlaşılabilir bir hal alması açısında sınıflandırılmaya ihtiyaç duymaktadır. Bu doğrultuda verinin daha şeffaf bir yapıda sunulabilmesi istenmektedir.

Bu bağlamda faktör analizinin temel amacı çok sayıdaki değişkenin daha az sayıda değişkene anlamlı faktörlerle indirgenmesi için uygulanan bir istatistiksel yöntemdir. Faktör analizinde faktörlere indirgeme ile yapışan sınıflandırmada faktörleşen değişkenler arasındaki ilişki göz önüne alarak yapılmaktadır. Bu bağlamda faktör analizindeki amaç, araştırmacıların veri setindeki karmaşıklığı azaltarak daha anlamlı ve az sayıda değişken ile sonuçlara ulaşma çabasıdır. [22]

2.3.1 Kavramsal Temeller

Çok değişkenli istatistiksel yöntem, birçok alanda farklı sebepler için kullanılabilmektedir. Bu yöntem ile birden fazla özelliğin kapsamlı analizine odaklanmaktadır. Bu hedefler arasında ilk olarak veri setlerinin anlaşılırlığını kolaylaştırmak açısından boyut indirgeme bulunmaktadır. Bir diğer kullanım amaçlarından bazıları ise değişkenlerin sınıflandırılması, korelasyonun ve kovaryansın değerlendirilmesi gibi nedenler bulunmaktadır. Bu hedefler doğrultusunda araştırmacı veri setinin karmaşık yapısından arınmış olacaktır ve kapsamlı bir bakış açısı sağlayabilecektir. [10], [23]

Faktör analizine dair bir bakış açısı kazanmak adına Şekil 2.20'de temsili bir görüntüsü mevcuttur. [24]



Şekil 2.20 Faktör analizinin görsel temsili

Uzun bir zaman dilimi boyunca uygulamaların sağlanması için yeterli hesaplamaların yapılamaması sebebiyle istatistiksel uygulamalar teorik potansiyelin gerisinde kalmıştır. Fakat yakın geçmişte teknolojinin ilerlemesi sayesinde çok fazla değişken içeren analizlerin hızlı, kolay ve istenilen hassasiyette gerçekleştirilebilmesi mümkün bir hale gelmiştir. Bu sebeple artık istatistiksel

analizlerin daha kapsamlı yapılabilmesi ve araştırmacıların ise veri setlerini daha etkin değerlendirmesine olanak sağlanmıştır. [23]

Faktör analizinin kökleri psikoloji bilimine dayanan ve sosyal tutum ve davranışların incelenmesine dayanan bir yöntemdir. Bu bağlamda bu yöntemin ilk incelenmesi 20. yüzyılın erken dönemlerinde Charles Spearman aracılığı ile Amerikan psikoloji dergisinde yayımlanan bir çalışması ile tanıtmıştır. Bu çalışma ile faktör analizinin temel oluşturmuş ve 1925 yılı itibariyle bu alandaki çalışmalar hızlıca artmıştır. [22]

1930 yılında ise Karl Person yaptığı çalışmalar ile faktör analizine temel eksen faktörü yöntemini kazandırmıştır. Bu bağlamda Person ile tasarlanmış olan bu yöntem Harold Hottelling tarafından 1933 yılında geliştirilmiştir. Ek olarak faktör analizine katkı sağlayan bir diğer bilim insanı ise Garnett'dir. 1919 yılında yapmış olduğu çalışmalar ışığında çoklu faktör analizi için büyük bir gelişme sağlamış olup akabinde Thurstone bu çalışmaları ve kendi çalışmalarıyla ilerletmiş ve daha etkin bir forma getirmiştir. [22], [25]

Faktör analizinin en yaygın kullanım amacı veri setindeki gizil yapıların ortaya çıkmasını hedeflemesidir ve bu amaç doğrultusunda sıklıkla kullanılmaktadır. Bu metodoloji ile ilgili olarak temel kavramlar ele alındığında korelasyon matrisi, öz değer, ortak faktör varyansı ve faktör yük değeri gibi kavramlar göz önünde bulundurulmaktadır. [26]

İlk olarak faktör analizinde oldukça önemli bir yere sahip olan korelasyon matrisi, gözlenen değişkenlerin arasındaki ilişkiyi ifade etmektedir. İkinci olarak öz değer de faktör analizinde oldukça sık karşımıza çıkan önemli bir kavramdır. Öz değer, her bir faktörün faktör yüklerinin karelerinin toplamını ifade etmektedir. Bu terim her bir faktör tarafından, açıklanan varyansın oranının hesaplanmasında ve faktör sayısının kararında kullanılan bir katsayıdır. Ek olarak öz değer bir veri setindeki değişikliğin ne kadar etkili bir şekilde açıklandığını ifade etmektedir. Bu bağlamda eğer ki öz değer yükselirse ilgili faktörün varyansının arttığı ifade edilmektedir. Başka bir ifadeyle veri setindeki değişkenliğin daha etkin açıklanabilmesi durumunun sağlanmasında yüksek öz değere sahip olması halindedir. Üçüncü kavram olarak ortak faktör varyansı da önemli bir kavramdır. Bu kavramın amacı varyansın farklı bileşenlerinin açıklanmasında kullanılmaktadır. Bu kapsamda üç

tane temel varyans vardır. Bunlar sırasıyla ortak varyans (ortak faktör varyansı), özgül varyans ve hata varyansıdır. Bir değişkende gözlenen varyansın temsili ortak varyanstır. Başka bir ifadeyle açıklanabilen varyans ortak varyans olarak ifade edilmektedir. Özgül varyans ise belirli bir değişkendeki gözlenmil toplam varyansı ifade etmektedir. Hata varyansı, veri setindeki açıklanamayan varyansı göstermektedir. Bu tanımlar doğrultusunda birbiri ile olan ilişkinin anlaşılabilirliği açısından ortak faktör varyansı ile özgül varyansın toplamı, faktör analizinde kullanılan modelin açıklanabilirliği ve güvenilirliğinin değerlendirilmesi noktasında oldukça etkilidir. Kısaca ortak faktör varyansının tanımı maddelerin faktörler ile olan çoklu korelasyonunun karesi ile ifade edilmektedir. Bu bağlamda ortak faktör varyansının yüksek çıkması modelin varyansının etkili bir şekilde açıklandığını ifade etmektedir. Bu durum ise analizin güvenilirliğinin arttırdığını göstermektedir. Son olarak faktör yük değeri kavramı ele alındığında bu kavram maddelerin faktörler arasında olan ilişkisini tespit eden bir katsayıdır. İstenilen durum faktör yüklerinin yüksek bir değerde olmasıdır. Eğer ki bir değişkenin faktör yükü 0,3 ise faktör tarafından açılanan varyans %9 olduğunu gösterir. Bu bağlamda genel olarak işaret ne olursa olsun 0.60 ve üzeri faktör yük değerlerini yüksek olarak ifade edebiliriz. Eğer ki 0.30-0.59 arasında bir faktör yük değeri varsa bu orta büyüklükte olarak kabul görmektedir ve bu durum değişken çıkartma durumunda bu nokta dikkate alınmaktadır. Bir başka açıdan faktör yüküne bakıldığında ise bu kavram sadece korelasyon değeri açısından incelenmesinden ziyade istatistiksel anlamlılık açısından da değerlendirilebilmektedir. Ancak bu noktada korelasyon değerinin düşük olması halinde, eğer ki örneklemin büyüklüğü artması halinde anlamlı çıkma olasılığının artabileceği göz önünde bulundurulmalıdır. Son olarak faktör yük değerinin bir diğer adı ise faktör katsayıdır. [26]

2.3.2 Faktör Analizi Çeşitleri

Bu başlık, açıklayıcı faktör analizi ve doğrulayıcı faktör analizi olmak üzere iki ana başlık altında ele alınmaktadır. Ardında ise açıklayıcı faktör analizi ve doğrulayıcı faktör analizinin benzerlik ve farklılıkları açısından incelenecektir.

2.3.2.1 Açıklayıcı Faktör Analizi

Açıklayıcı faktör analizinin temel tanımı değişkenler arasındaki doğrusal yapıyı belirterek boyut indirgenmesini hedeflemektedir. Belirlenmiş değişkenlerin

arasındaki ilişkiyi bir araştırmacı bu yöntem ile temellendirebilir. Bu yöntem özellikle ölçek geliştirme başta olmak üzere güvenilirlikte de sıklıkla tercih edilmektedir. [22]

2.3.2.2 Doğrulayıcı Faktör Analizi

Doğrulayıcı faktör analizinin temel tanımı temelinde geliştirilmiş olan bir ölçme aracı aracılığıyla elde edilmiş olan veriler ışığında modeldeki yapının doğrulanıp doğrulanmadığını test etmektir. Bu yöntem özellikle mevcutta olan teorik yapının geçerli olup olmadığının belirlenmesinde oldukça sık tercih edilmektedir. Kısacası bu yöntem aslında önceden belirlenmiş olan bir model üzerinde odaklanarak belirlenen bu modelin veriye olan uygunluğunun test edilmesini hedeflemektir. Bu bağlamda bu yöntemin kullanılabilirliği açısında araştırmacının açıklayıcı faktör analizinin de biliniyor olmasında fayda olacaktır çünkü doğrulayıcı faktör analizinin temelini oluşturan yapı açıklayıcı faktör analizidir. [22]

2.3.2.3 Benzerlik ve Farklılıklar

Açıklayıcı faktör analizi (AFA) ile doğrulayıcı faktör analizi (DFA) arasında hem benzerlik hem de farklılık açısından çeşitli durumlar gözlemlenmektedir. Benzerlikleri açısından bakıldığında her iki modelinde temelinde doğrusal istatistik modeli olması, normal dağılıma uygunluğu ve modellerin gizil yapıları göstermesi yer almaktadır. Açıklayıcı faktör analiziile doğrulayıcı faktör analizi arasındaki farklılıklar açısından bakıldığında ise DFA'da önceden belirli olan ve faktör sayısı belirli bir modelde çalışılır, ek oarak mevcut olan modelin doğrulanmasını ya da test edilmesinini hedeflerken, AFA'da ise varyansı maksimum tutarak modeldeki yapıların ortaya çıkmasını hedefler ve yeni bir modelin keşfedilmesine odaklanan analitik bir süreci göstermektedir. Bu temel farklılıklar, bu iki yöntemin uygulama alanlarının belirlenmesini sağlayan önemli unsurları göstermektedir. [22]

2.3.3 Faktör Analizi Varsayımları

Faktör analizi için oldukça önemli olan bir nokta değişkenlerin normal dağılıma sahip olmasıdır. Bu durum tüm lineer kombinasyonlar için uygundur.[27]

Faktör analizinde değişkenlerin arasındaki ilşikinin lineer olduğu varsayımı altında, model de doğrusal bir model olduğundan dolayı analiz edilecek olan değişkenler genellikle eşit aralıklı ölçme düzeylerinde ölçülmesi beklenir. Bu durum sebebiyle

değişkenlerin ağırlıkları eşit tutulmasına olanak sağlar. Bunlara ek olarak ise istatistiksel anlamda birbirleri arasında bir ilişkisinin olmadığı bir değişkenin, korelasyon matrisine bakarak bu değişkenin analizden çıkarılmasına olanak sağlamaktadır. [28]

Faktör analizine sokulacak olan verinin metrik ölçülmesi istenilen bir durumdur. Özellikle oran ölçekli değişkenlerin bulunduğu analizde daha iyi sonuçlar elde edilir. Bu bilgiye ek olarak aralık ölçekli değişkenlerin bulunduğu analizde oran ölçekli kadar olmasa da kısmen etkili sonuçlara ulaşılabilmesini sağlayabilmektedir. Eğer ki faktör analizinde metrik olmayan ölçek olan sıralı ölçekle ölçülecek olursa, Likert, Thurstone ya da Goodman ölçeklerinin kullanılması gerekmektedir. Bu bilgilere ek olarak eğer ki veri seti çok fazla sıralı ölçek ve ikili ölçek değişken bulunduruyorsa, bu durumda yapılan analizin sonuçlarındaki faktörlerin yorumlanması zor hale gelir. Bu bağlamda, analize çok düşük ya da çok yükse korelasyona sahip veriler analize dahil edilmeyebilir. [8]

2.3.4 Faktör Analizi Aşamaları

Faktör analizi, çok sayıda aşamaya sahip analiz yöntemidir. Bu aşamalar sırasıyla problemi tanıma ve veri toplama, korelasyon matrisinin oluşturulması, faktör sayısının belirlenmesi, faktör eksen döndürme olarak tanımlanmaktadır. [26]

İlk olarak problem tanımı ve veri toplama aşamasında amaç faktör analizinin uygulanması için gerekli hazırlıkların yapıldığı ilk adımdır. Bu aşamada, ilk hedef kullanılacak olan değişkenler için teorik temeli ve literatür araştırması yapılır. Bunlara ek olarak araştırmacı tarafından yapılan çalışmalar ile ölçülme uygunluğunun değerlendirilmesi yapılır. İkinci aşamada ise korelasyon matrisinin oluşturuşmasıdır. Bu evre aslında faktör analizinin başlangıcını belirtmektedir. Bu bağlamda analiz için seçilmiş olan değişkenlerin arasındaki bağlantının anlaşılması için bu matristen faydalanılmaktadır. Bu aşamada korelasyon matrisinin amacının değişkenler arasındaki ilişkinin anlaşılması sağlar ve bu analiz için kullanılacak veri setini sunmaktadır. Üçüncü adımda ise korelasyon matrisinin oluşmasının ardından faktör sayısının belirlenmesidir. Bu adımdan önce hazırlanmış olan korelasyon matrisine bakarak uygun olan faktör çıkarma yöntemi seçimi ve başlangıç çözümünün oluşturulmasının sağlanmasını içermektedir. Son aşama ise faktör eksen döndürmü aşamasıdır. Bu adımda istenilen durum, başlangıç faktör

matrisininin elde edilmesinin ardından ortaya çıkan faktörlerin isimlendirimesinin ve kolay yorumlanması amacı ile faktörlerin temsili olan eksenler üzerinde eksen kaydırma yapabilmesidir. [26]

Sonuç olarak faktör analizi ile elde edilmiş olan birincil temel bileşen, veri setindeki varyansın büyük çoğunluğunu açıklamasından ve farklı değişkenlerin lineer kombinasyonlarından oluşması sebebiyle değişkenlerin birçoğu üzerinde geniş pozitif yüklemeye sahip olduğu için bu genel faktör şeklinde isimlendirilmektedir. Bu bağlamda birincil temel bileşen aslında genel bir faktörün temsilidir. Ancak uygun olmamakla birlikte birincil temel bileşenin varlığı aslında genel bir faktörün varlığının göstergesi olarak gösterilebilmektedir. Akabinde diğer faktörler ise hem pozitif hem de negatif yüklemeler sahip olan kutuplu faktörlerdir.

2.3.5 Verinin Uygunluğu

Faktör analizine başlamadan önceki adımda sonuçların güvenilirliğinin ve doğru bir şekilde yorum yapılabilmesi amacıyla veri setinin bir takım belli özelliklere sahip olması gerekmektedir. Bu noktada, normallik, çoklu doğrusal bağlantı ve doğrusallık gibi varsayımları veri setinin sağlaması gerekmektedir. Faktör analizinde likert ölçek gibi verinin düzgün ölçülmesini sağlayacak bir ölçek kullanılması tavsiye edilmektedir. Bu sayede, araştırmacı analiz sürecinde ulaştığı sonuçların güvenilir ve geçerli bir sonuç elde edebilmesi sağlanacaktır.

Faktör analizinde temel hedef değişkenler arasındaki ilişkinin anlaşılması ve değişken sayısını azaltarak az sayıda temel faktör elde etmektir. Bu bağlamda doğru sonuçlara ulaşabilmek adına belli başlı önemli etkenlere dikkat edilmelidir. [28]

Bunlardan ilki faktör analizi için gerekli gözlem sayısı 100 veya daha fazla olması önerilmektedir. Bilimsel araştırmalarda araştırmacıların faktör analizi yaparken gözlem sayısı 50 veya daha altı ise bu analiz için gerekli görülmemekte ve faktör analizi önerilmemektedir. Bunu sebebinin analizin güvenilirliğinin azalacağı ve elde edilen sonuçların belirsizlik içermesine neden olacağıdır. Bu bağlamda analize başlamadan önce veri setinin büyüklüğü ve analize uygunluğunun değerlendirilmesi önemlidir. Bu sayede güvenilir ve anlamlı sonuçları elde etmek için öncesinde bu aşamalara dikkat edilmelidir.

İkinci olarak değişkenler arasındaki korelasyon seviyesinin yeterli olması gerekmektedir. Çünkü analizin sonuçlarında anlamlı faktörler oluşması için, değişkenlerin arasındaki yüksek korelasyonla ilişkilidir. Bu bağlamda kısmi korelasyon katsayısı incelendiğinde, bu değerin 0,30'dan büyük olması beklenmektedir. Bu sayede analizin sağlam temele dayanır ve güvenilirliğini arttırmaktadır. [8]

Bunlara ek olarak bir veri setinin faktör analizine uygun olup olmadığının ölçülmesi amacıyla Bartlett testi ve Kaiser-Meyer-Olkin (KMO) ölçütlerine başvurulması önem arz etmektedir. Bu bağlamda Bartlett testi, değişkenlerin arasındaki korelasyon seviyesinin anlamlılığını değerlendirirken KMO ölçütü ise faktör analizine uygun olup olmadığını test etmektedir. Bu ölçütler sayesinde faktör analizine başlamadan önce veri setinin uygunluğunun ölçülmesi için önem arz eder.

2.3.5.1 Bartlett Testi

Bartlett Testinin diğer adı olan küresellik testinin amacı korelasyon matrisinin birim matrise eşitliğini sınamaktır. Eğer ki korelasyon matrisi birim matrise eşit ise veri seti için faktör analizine uygunluğu kabul edilmez. Ancak korelasyon matrisi ile birim matris eşit değil ise veri setinin faktör analizi için uygunluğu kabul edilir. Bu bağlamda bu testin temelindeki amaç veri setinin faktör analizi öncesinde faktörlendirmeye uygun olup olmamasının yorumlanmasında önem arz etmektedir. [29]

Faktör analizi işlemine geçilmeden önce, bu test oldukça kritik bir adımdır. Bunun sebebinde ise eğer ki bartlett testi veri seti için istatistiksel anlamda anlamlı çıkarsa, faktör analizi sürecine geçiş sağlanacaktır. Küresellik testinin öneminin daha iyi anlaşılması için bir benzetme yapmak gerekirse, varyans analizinde varyansların homojenliğinin kritik bir aşama olması gibi bartlett testide faktör analizi için kritik bir aşama olduğu söylenmektedir.

Özetlemek gerekirse bartlett testi bir veri setinde değişkenlere ait olan korelasyon matrisinin anlamlılığının test edilmesinde öne çıkar.[10]

2.3.5.2 Kaiser-Meyer- Olkin (KMO) Testi

Kaiser-Meyer-Olkin ölçütü, faktör analizi için veri setinin uygunluğunun test edilmesini sağlamaktadır. Bu bağlamda bu testin amacı korelasyon katsayılarıyla

kısmi korelasyon katsayılarının arasındaki büyüklük farkının incelenmesi ve bu sayede faktör analizine olan uygunluğunun değerlendirilmesini hedeflemektedir. [30]

KMO ölçütü ile barlett testinin farkı ise KMO bir veri setinin faktör analizi için uygunluğunu ölçen bir araç olmasıdır. [29]

KMO uygunluk testi için önerilen aralık ve yorumları değerlendirmek gerekirse sırasıyla eğer ki KMO ölçüsü 0,90 veya daha üstü bir değer çıkması mükemmel, 0,80 – 0,89 arasında bir değer çıkması çok iyi, 0,70 – 0,79 arasında çıkması iyi, 0,60 – 0,69 arasında çıkması orta, 0,50 – 0,59 arasında çıkması zayıf ve 0,50 değerinin altında çıkması ise zayıf olarak nitelendirilmektedir. Eğer ki KMO ölçütü 0,50'in altında çıkması durumunda bu veri setinin faktör analizine uygun olmadığını göstermektedir. Böylece bu değerler sayesinde KMO değerinin aslında faktör analizi için uygunluğunun değerlendirilmesinde kullanılmaktadır. [29]

Bir veri setinin faktör analizine uygunluğu bakılırken istenilen KMO ölçütünün 0,50'den büyük bir değere sahip olmasıdır, ancak 0,80'den büyük olan bir KMO değeri faktör analizi için sağlam sonuçların elde edilebileceğini göstermektedir. [31]

Bu bilgilere ek olarak korelasyon matrisinin negatif çıkması halinde KMO değeri ölçülmez. Bu durumla karşılaşıldığında faktör analizine uygunluğun ölçülmesi için başka bir yol izlenir. Bu yöntem ortak varyansa bakılması durumudur. Bu bağlamda değişkenlerin analize olan uygunluğunu ortak varyans değerleri göstermektedir. Eğer ki ortak varyans değerleri 0,50'den büyük ise veri setinin faktör analizi için uygun olduğu söylenebilmektedir. [29]

2.3.6 Faktörlerin Belirlenmesi

Bu başlık altında temel bileşen analizi ve maksimum olabilirlik yöntemleri incelenecektir. Bu yöntemler arasında genel kabul görmüş olan ve sık sık uygulanan yöntem, temel bileşenler analizi yöntemidir.

2.3.6.1 Temel Bileşen Analizi (TBA)

Temel bileşenler analizi, 1901 yılında Karl Pearson tarafından öne sürülmüş ve ardından 1933 yılında ise Hotelling tarafından ise geliştirilmiş bir analiz

yöntemidir. Bu analiz yöntemi çok değişkenli analizlerden en yaygın olanı ve en eski olanıdır. [32]

Bu yöntemin temelinde değişkenler arasındaki ilişkinin anlaşılması için birkaç doğrusal kombinasyon kullanma işlemidir. Daha kapsamlı ele almak gerekirse bu yöntemin asıl hedefi birbirleri ile ilgili olmayan yeni değişkenler elde edip böylelikle değişken sayısının azalmasını sağlayıp yeni bir değişkenler kümesi oluşturmaktır. Bu bağlamda yeni değişkenler kümesi ile belli başlı analiz yöntemlerine uygun bir hale getirilmiş olacaktır. [33]

Temel bileşen analizi, literatürde kimi zaman faktör analizi kimi zaman temel faktör analizi kimi zaman da ortak faktör analizi olarak isimlendirilen bu metodolojinin temel amacı veri setinde maksimum sayıda ortogonal varyansı çıkarmaktır. Bu yöntemin birçok avantajı vardır. Bunlardan birisi, bu yöntem sayesinde hata varyansının kaldırılmasıyla ortak varyansın belirlenmesidir. Bu yöntem ortak varyansa odaklanmaktadır. Bu durum aslında faktör analizi yöntemine uygun bir yaklaşım sunmaktadır. Ortak faktör ile ortak varyansın tanımları ise, ortak faktör değişkenleri etkileyen ve değişkenlerin arasındaki ilişkiyi açıklayan yapı olarak tanımlanmaktadır. Ortak varyans, ortak faktörler ile açıklanan varyans olarak tanımlanmaktadır. Fakat bu yöntemin faktör bulma teknikleri ile kıyaslandığında korelasyonu yeniden oluşturulması aşamasında zaman zaman başarılı sonuçlara ulaşmayabilmektedir. [29]

Bu bilgilere ek olarak bu analiz yönteminin sıklıkla tercih edilme nedenleri arasında hata varyansının düşük olması, az sayıda faktör ile büyük bir oranda açıklanabilmesi ve elde edilmiş olan sonuçların başka analiz yöntemlerinde de kullanılabilir olmasıdır. [34]

Temel bileşen analizinde p sayıda değişken, birbirinden bağımsız olan p sayıdaki temel bileşen ile temsil edilmektedir. Bu bağlamda bu yöntemde birinci temel bileşende en çok katkıya sahipken diğerlerinde giderek azalarak toplam varyansa dahil olmaktadır. Kısacası temel bileşen analizi, birinci temel bileşen itibariyle sırayla toplam varyansa rn yüksek seviyeden katkıda bulunur ve giderek azalan bir eğilim göstermektedir. En genel kapsamda başlangıçtaki değişken sayısı ile ortak faktör sayısı birbirlerine eşit iken sonrasında elde edilen ortak faktörlerin birkaçı toplam varyansın büyük ve önemlş bir çoğunluğunu açıkladığını varsayarken diğer

faktörlerin daha az açıklanabilir olduğunu yani bu ortak faktörlerin düşük varyanslara sahip olduğunu varsayar. [35]

2.3.6.2 Maksimum Olabilirlik Yöntemi

Maksimum olabilirlik yöntemi, doğrulayıcı faktör analizinde kullanılması tercih edilen bir yöntemdir. Bu yötemin avantajları sırasıyla bir modelin uyumluluk indeksinin hesaplanmasındaki kullanılabilirliği ve araştırmacının araştırma başında elde ettiği faktörler faktör sayısının belirlenmesini, kurulmuş olan hipotezin test edilebilirliği ve faktörlere anlamlılık testinin uygulaması gibi becerileri mevcuttur. [22]

Bu yöntemin uygulanabiliriliğinin en elverişli olduğu veri seti, çoklu normallik varsayımının geçerli olduğu bir veri setidir. Çoklu normallik varsayımının karşılandığı modeldeki veri setine dayalı korelasyon katsayısı değerlerinin alabileceği maksimum değere ulaşmayı hedeflemektedir.

2.3.7 Faktör Sayısının Belirlenmesi

Faktör sayısının belirlenmesi, veri setindeki değişkenlerin daha az sayıda faktör elde edilmesini hedefleyen bir adımdır. Bu süreçte değişkenlerin arasındaki korelasyonun en uygun ve iyi şekilde temsil eden faktörlere ulaşmak asıl amaçtır. Bu amaç doğrultusunda çeşitli yöntemler mevcuttur. Bu yöntemler sırasyla yamaç grafiğine göre belirleme, özdeğere göre belirleme ve varyans oranına göre belirleme şeklinde sıralanabilir.

2.3.7.1 Scree Test (Yamaç Grafiği) Kriterine Göre Belirleme

Faktör sayısının scree test kriterine göre belirlenmesi yönteminde temel hedef özdeğerlerin grafiğinin çizilmesi ve ardında yatay ve dikey eksendeki kombinasyonlarına göre faktör sayısının belirlenmesidir.

Scree test kriterine göre yapılan faktör belirleme yöntemi en sık kullanılan yöntemlerden birisidir. Bu doğrultuda oluşan her bir faktör için ilgili özdeğerin yorumlanması için bir gösterim sağlamaktadır. Yamaç grafiğine göre faktör sayısının belirlenmesi aşamasında incelenilen nokta grafikte belirgin bir ölçüde düşüş yaşadığı doğrultuya kadar olan kısımdaki faktörler faktör sayısını belirlemektedir. Bu noktadan sonra olan faktörlerin etkisi oldukça düşüktür.

Etkisinin az olduğu faktörler grafikte adeta bir yatay doğrultudadır. Özetle bu yöntem ile elde edilen sonuçların dengeli ve anlamlı bir şekilde belirlenmesine olanak sağlamaktadır.

2.3.7.2 Özdeğere Göre Belirlme

Faktör sayısının özdeğere göre belirlenmesi yönteminde amaç ilk olarak faktör sayısı kadar özdeğer hesaplanmasıdır. Bu özdeğerler hesaplandığında faktör yüklerinin kareleri toplamı ile faktör sayısı belirlenmektedir. Bir faktörün açıklayabildiği toplam varyansa özdeğer denmektedir. Eğer ki bir faktörün özdeğer istatistiği 1'den büyük ise o faktör kabul edilir. Bir başka kriter ise 0.7 ve üzerinde olan özdeğerlerin modele dahil edilmesini önermektedir. Bu Joliffe kriteri olarak adlandırılmaktadır. Bu kriterdeki temel hedef modele dahil edilmesi için gerekli olan minimum özdeğer seviyesini belirler. [30]

2.3.7.3 Varyansın Oranına Göre Belirlme

Faktör yapısındaki belirginlik, analizdeki varyans oranına bağladır. Bu oran genellikle %40 ile %60 olarak kabul görmektedir. Bu orana sahip bir analizin sonuçlarının yeterli bir güçte olduğu kabul edilmektedir. En genel kapsamda literatürdeki araştırmalar ışığında varyans oranının %50 ve üstü bir oranın yapılan analiz için yeterli olduğu kabul edilmektedir.

2.3.8 Faktörin İsimlendirilmesi

Faktör analizi yöntemi ile elde edilmiş olan yeni faktörlerin tanımlanması kimi zaman zorlu olabilmektedir. Oluşan yeni faktörde birbiri ile düşük korelasyona sahip olan değişkenler aynı faktör altında birleşebilmektedir. Böyle durumlarda faktörün isimlendirilmesinde dikkate alınacak olan husus en yüksek faktör yüküne sahip olan değişken dikkate alınarak yapılır. Başka bir ifadeyle faktör yük değerine göre faktörlerin isimlendirilmesi yapılmaktadır. Bu isimlendirme işleminde en yüksek faktör yükünün değerine bağlı olarak çeşitli ölçeklere dayanmaktadır. Bu bağlamda faktörlerin isimlendirme işlemi yapılırken dikkat edilecek olan hususlardan vardır. [36]

Faktörlerin isimlendirimesi yapılırken ilk olarak eğer ki faktör yükü 0.7'den düşük ise bu değişkenlerin yorumlanmasında özenli olunmalıdır. Böyle durumlarda ise gözlenen değişkenin faktör yükü, faktör yükünün büyüklüğüne bağlı olarak daha az açıklanabilir. İkinci bir husus olarak ise iki veya ikiden daha çok anlamlı yük

taşımakta olan faktörler ise genel olarak ortak faktör olarak isimlendirilmektedir. Bir diğer husus ise faktör yükünün en yüksek değerde olan o faktör temsilci olarak kabul edilmektedir. Bu bilgiye ek olarak ise faktör yükünün yüksek bir değerde olan bir ya da birden fazla değişkenin benzer olan özellikleri dikkate alınarak föktere isimlendirme yapılabilir. Başka bir husus olarak pozitif ve negatif faktör yüklerine sahip olan zıt işaretli değişkenler ayrı gruplara ayrılabilir. Başka bir durum ise eğer ki bazı değişkenlerin faktör yükü düşük ise bu değişkenlerin yapılan analizden çıkartılmsında bir sakınca olmayacaktır. Çıkarılan değişkenlerden sonra ise analiz tekrarlanıp yeniden bir analiz yapılabilir. [36]

2.3.9 Faktör Döndürmesi

Faktör analizinde elde edilen unsurların; bağımsızlık, netlik ve anlamlılık gibi belli başlı ölçütlere uymasını amaçlayan bir araştırmacı, bu unsurları faktör rotasyonuna tabi tutabilmektedir. Faktör döndürmeye tabi tutulan bu unsurların temel matematiksel özellikleri değişmemektedir. Eksenlerin döndürülmesi yapıldığında bir maddenin belli bir unsurundaki (faktörler) yükler artar iken, bir diğer unsurundaki yüklerin azaldığı görülmektedir. Kısacası faktör rotasyonu sayesinde yüksek ilişkili maddelerin bulunması kolaylaşır, ve bu sayede unsurlar yorumlanmaya daha açık hala gelmektedir. [9]

Thurstone tarafından 1947 yılında belirttiği ve aşağıda açıklanacak basit yapının elde edilmesi amacıyla faktör analizinin sonuçlarının daha da anlamlandırılması amaçlanmaktadır.

İlk husus faktör döndürmesi yapabilmek için her bir satırın veya faktör yükleri matrisindeki yüklerden en az bir tanesinin sıfır değerinde olması gerekmektedir. Bir diğer unsur ise faktör sayısını k olarak belirlendiğinde, her bir faktörde minimum k sayısı kadar sıfır değeri bulunmak zorundadır. Ayrıca yüklerin, yükler matrisinin her bir sütun çiftinin birinde bulunması durumunda diğerinde bulunmaması gerekmektedir. Bu bilgilere ek olarak faktörün dört ya da daha fazla olduğu durumlarda, faktör çiftleri için değişkenlerin çoğunun sıfır yüküne sahip olması gerekmektedir. Ayrıca, eğer ki föktör sayısı dört yada daha fazla ise, her bir faktör çifti için yük matrisinde az sayıda değişkenin yük değeri olmalıdır. [9]

Faktör analizinde, iki çeşit rotasyon yaklaşımı kullanılmaktadır. Faktör döndürme için, dik ve eğik olmak üzere iki farklı kategoriye ayrılmaktadır.

2.3.9.1 Dik (Orthogonal) Rotasyon

Söz konusu dik faktör rotasyonu olduğunda rotasyon, eksen konumları değiştirilmeden yapılmaktadır. Başka bir ifadeyle belirtmek gerekirse, faktöre 90 derecelik bir rotasyon işlemi yapılır. Dik faktör rotasyonu ile elde edilmiş olan faktörler birbirinden bağımsız olmaktadır. [30]

Dik faktör rotasyonu yapıldıktan sonra, faktör yükleri matrisindeki değerler ile faktör değişkenleri arasındaki ilişki incelenmektedir. Bu inceleme yapılırken araştırmacının yalnızca yük değerleri 0.32'den daha yüksek olan değerlere odaklanmalıdır. Yük değerleri %32 ile %45 arasında olan değerlerin zayıf kabul edilir ve böyle kabul edilmesinin nedeni ise varyansın yalnızca %10'unu kapsamasından dolayıdır. Varyansın %20'lik kısmını kapsayan yani yük değerleri %45 ile %55 arasında olanlar orta düzey, varyansın %30'unu kapsayan yani yük değerleri %55 ile %63 arasında olanlar iyi düzey, varyansın %40'ını kapsayan yani yük değerleri %63 ile %71 arasında olanlar çok iyi düzey ve son olarak varyansın %50'sini kapsayan yani yük değerleri %71'den fazla olanlar mükemmel düzey olarak sınıflandırılmıştır. [37]

Dik faktör döndürmesi yapmak için genellikle Varimax, Quartimax ve Equamax teknikleri kullanılmaktadır. Bunlardan ilki, Kaiser tarafından geliştirilmiş olan dik faktör rotasyonu Varimax yani Maksimum Değişkenlik olarak adlandırılmaktadır. Bu tekniğin amacı, faktör varyanslarını maksimize etmek ve daha açık, basit bir yapı haline getirebilmektir. Varimax tekniğinde faktör matrisindeki sütunlar üzerinde çalışılır. Her satırda bulunan belirli yük değerleri 1'e yaklaştırılırken, diğer yük değerleri 0'a yaklaştırılmaktadır. Bahsettiğimiz Varimax tekniği, dik faktör rotasyonları arasında en çok tercih edilen yöntemlerdendir. Bir diğer yöntem olan quartimax tekniğinde ise faktör matrisindeki satırlar üzerinde çalışılarak, daha basit bir yapı elde edilmek amaçlanmaktadır. Bu teknik, iki faktörlü modeller için geliştirilmiş ve en iyi sonuçların elde edildiği dik faktör rotasyonu tekniğidir. Bu yöntemde, her satırda bulunan belirli yük değerleri büyütülür ve 1 değerine yaklaştırılır iken, diğer yük değerleri küçültülür ve 0 değerine yaklaştırılır. Varimax ile quartimax teknikleri arasındaki fark ise varimax tekniğinde gerçekleştirilen işlem faktör merkezinli odaklı iken, quartimax tekniğinde değişken merkezine odaklıdır. Tekniklerden bir diğeri ise equamax yöntemidir. Equamax tekniğinin, varimax ile quartimax yöntemlerinin kesişme noktası olduğunu söylenebilmektedir.

Bu teknikte faktörler ve değişkenlerin her biri kullanılarak işlem yapılır ve bu şekilde daha basit bir yapı elde edilmek amaçlanır. Fakat bu teknik araştırmacılar tarafından, diğer iki teknik kadar kabul edilmemiştir. Bu yüzden Equamax tekniği nadiren kullanılmaktadır. [37]

Dik faktör rotasyonunun temel özellikleri ise ilk olarak, dik faktör rotasyonu yapıldığı takdirde temel bileşenler arasında istatiksel bir bağımsızlık sağlanmaktadır. Başka bir deyişle, rotasyon yapılan faktörlerin arasındaki açının cosinüsü 0 haline getirilmiş olur. İkinci olarak, faktörlerin rotasyonu yapılmadan önceki ortak varyansı ile rotasyon yapıldıktan sonraki ortak varyansı eşdeğer olmaktadır. Fakat dik faktör rotasyonu sonrasında, faktör değişkeninin açıklanan varyansı farklılık göstermektedir. Son olarak, dik faktör rotasyonu sonrasında elde edeceğimiz faktör sıralaması ile rotasyon öncesindeki faktör sıralamasının birbirinden farklı olacağını söylenebilmektedir. [33]

2.3.9.2 Eğik (Oblique) Rotasyon

Eğik faktör rotasyonu teknikleri, dik faktör rotasyonu tekniklerine kıyasla daha çok tercih edilmektedir. Bunun sebebi, eğik faktör rotasyonu tekniklerinin daha üstün sonuçlar ortaya koyabilmesidir. Oblik rotasyon olarak da adlandırılan, eğik faktör rotasyonunda eksenlerin dik olma koşulu aranmamaktadır. Her bir faktöre birbirinden bağımsız olarak döndürme yaptırılmaktadır. Eksenlerin dik olmaması nedeniyle eğik faktör rotasyonu yapılmasına karar verildiğinde, faktör yüklerinin yorumlanması için kullanılabilecek iki temel yöntem bulunmaktadır. Bu iki temel yöntemin ilkinde, değişkenlerin yerine geçen her noktanın rotasyon yapılmış eksenler üzerindeki izdüşümleri ile eksenlere paralel olan doğrular kullanılarak hesap yapılmaktadır. Elde ettiğimiz bu doğrular ‘pattern loadings’ yani örüntü yükleri olarak adlandırılmaktadır. Yöntemlerin ikincisinde ise, değişkenlerin yerine geçen her noktanın eksenler üzerindeki izdüşümleri ile eksenlere dik olan doğrular kullanılarak hesap yapılmaktadır. Bu şekilde hesaplandığında, rotasyon yapılmış eksenler üzerindeki yük değerleri ‘structure loadings’ yani yapı yükleri olarak adlandırılmaktadır. Bahsedilen bu yapı yükleri, orijinal değişkenler ile faktörler arasında gerçek ilişkiyi gösteren katsayılardır. [7]

Eğik faktör döndürme teknikleri arasında Oblimax, Quartimin, Covarimin, Biquartimin, Oblimin ve Binoramin adlı teknikler en çok tercih edilen tekniklerdir.

Eğik faktör rotasyonunda bahsedilen bu altı tekniğin incelenmesi sırasıyla ilk olarak Oblimax tekniği, 1961 yılında Saunders tarafından geliştirilmiştir. Bu yöntemde kurtosis yani basıklık katsayısını maksimize etmek temel amacıdır. Oblimax tekniğinin asıl amacının bu olduğu kabul edilmektedir. Kullanılan tekniklerden bir diğeri ise Quartimini tekniğidir. Quartimini tekniği ise 1953 yılında Carroll tarafından geliştirilmiştir. Faktör yüklerinin kareleri çarpımının toplamını minimum seviyede tutmaya çalışmak, tekniğin asıl hedefidir. Diğer bir teknik olarak ise Covarimin tekniğinden bahsedilebilir. 1953 yılında, yine Carroll tarafından geliştirilen Covarimin tekniğinin hedefi ise fonksiyonu minimize etmektir. En çok kullanılan teknikler arasında, son olarak Biquartimin tekniğinden bahsedilebilir. Bu yöntemde ise Quartimin ve Covarimin yöntemlerinde kullanılan fonksiyonları içerir. Bu teknikte ifade edilen Biquartimin fonksiyonunun minimum olması hedeflenir. [10]

Eğik faktör rotasyonunun temel özellikleri şu şekilde sıralanabilmektedir: Korelasyon matrisinin faktörleri birbirine dik ise, yük değerleri -1 ile +1 arasında olmak zorundadır. Eğik faktör rotasyonunda ise, istisnai olarak bazı mutlak yük değerleri 1'den büyük olabilmektedir. Faktör skorları birbirleri ile ilişkili olmak zorundadır. Yine eğik faktör rotasyonunda, faktörlerin açıklanan varyans yüzdesi veya değişkenin ortak varyansı, yük değerlerinin karelerinin toplamından hesaplanamamaktadır. Structure loadings yani faktör yapı yükleri ile pattern loadings yani örüntü yükleri matrisi arasında bariz olarak görülebilen bir fark bulunmaktadır. [33]

2.3.10 Faktör Skorları

Faktör skorları, asıl değişkenlerin yerine geçebilecek, daha az boyutlu ve anlamlı değişkenler sunmaktadır. Faktör skorları sayesinde, analizin sonuçları daha anlamlı hale getirilmekte ve analizdeki değişkenler daha mantıklı ve etkili bir şekilde ifade edilmektedir. Her bir faktörün hesaplanmış skorlarının tahminleri faktör skorları olarak isimlendirilmiştir. Faktör skorları olarak isimlendirilen bu tahminlerin yapılacağı süreçte, faktör skorları çeşitli karmaşık istatistiksel yöntemler ile hesaplanmaktadır. Hesaplanan faktör skorları sayesinde yapılan analizin anlamı derinleştirilir. Faktör değerleri çok daha etkili ve kolay bir şekilde anlaşılabilir hale getirilmiş olmaktadır. Faktör skorları, bilgisayar paket programlarında oldukça

yayın bir şekilde kullanılmaktadır. Bahsedilen faktör skorlarını tahmin etme yöntemleri ise aşağıdaki gibi sıralanabilmektedir. [7]

Faktör skorlarını tahmin edebilmek için kullanılan yöntemlerden biri Bartlett yöntemidir. Bu yöntemi, 1950 yılında Bartlett tarafından geliştirilmiştir. Bu yöntemin amacı, ortak faktörlerinin p-değişkeni üzerinden ve özgür faktör katkılarının n-bireyi üzerinden minimum hale getirilmesidir. Bu yöntemde faktör skorları tarafsızdır. Faktör skorları yalnızca kendi faktörleri ile bağlantılıdır. Kullanılan bir diğer yöntem ise Thomson yöntemidir. Thomson yöntemi, 1939 yılında yönteme ismini veren Thomson tarafından önerilmiş bir yöntemdir. Bu yöntemin Bartlett yöntemi ile belli başlı benzerlikleri bulunmasına rağmen temelinde önemli farklılıklar bulunmaktadır. Bartlett yönteminde her bireydeki değişim ortak faktörler dikkate alınarak hesaplanırken, Thomson yönetiminde regresyon kuralı kullanılarak hesaplama yapılmaktadır. Regresyon kuralı ile her değişkenin değişim oranı en aza indirgenir ve n-bireyinin her kümesindeki faktörlerin değerleri tahmin edilir. Faktör skorlarının tahmininde Anderson ve Rubin yöntemi de kullanılmaktadır. Bu yöntem 1956 yılında, Anderson ve Rubin tarafından geliştirilmiştir. Bu yöntemin asıl amacı, birbirinden bağımsız faktör skorları bulabilmektir. Bir faktör skorunun ortalaması 0, standart sapması ise 1 olarak standardize edilmektedir. [24], [38]

2.3.11 Faktörin Yorumlanması

Faktör analizinin uygulanması belirli istatistiksel kısıtlar üzerine kurulmuştur. Bu sebeple araştırmacıların bunları göz önüne alarak dikkatli bir şekilde gerçekleştirmesi gerekmektedir.

Faktör analiziyle birlikte veri setindeki çok sayıdaki değişkenin daha az sayıda ve anlamlı değişkenlere dönüştürülme sürecidir. Bu bağlamda oluşmuş olan yeni değişkenlerin, içlerindeki değişkenleri özetleyici niteliğindedir. Bu sayede benzer özellikteki birçok değişken anlamlı bir bütün halinde temel bir faktöre dönüşmektedir. Bu sayede faktör analizi aslında karmaşık yapılı bir halden kurtulmak, bu sayede daha iyi anlamak, ölçmek ve değerlendirmek amacıyla araştırmacıların önemli bir araç olarak gördüğü bir yöntemdir. [39]

2.3.12 Faktör Analizi Uygulamalarında Karşılaşılan Sorunlar

Bu bölümde, faktör analizi yönteminin uygulanma aşamasında araştırmacıların karşılaşabileceği sorunlar ele alınacaktır. Bunlardan başlıcaları örneklem büyüklüğü, normallik ve doğrusallık açısından ortaya çıkabilecek potansiyel sorunlardır.

İlk olarak örneklem büyüklüğü incelenecek olursa eğer ki küçük bir örneklemden elde edilen ilişki katsayıları var ise bu daha az güvenilirlik göstergesi olarak kabul edilmektedir. Bu bağlamda ilişkilerin güvenilirliği için örneklem büyüklüğünün uygun bir seviyede olması kritik bir unsurdur. İkinci olarak normallik açısından incelendiğinde araştırmacının veri setinin normal dağımda olması istenmektedir. Bu durum faktör analizinde güvenilir sonuçlar elde edebilmek adına önemlidir. Bu nedenle eğer normal dağılım söz konusu değilse güvenilir sonuçlara ulaşılamayabilir. Ancak normallik varsayımına uymayan durumları düzeltmek adına birçok yöntem mevcuttur. Üçüncü olarak doğrusallık incelendiğinde ise faktör analizi öncesinde doğrusallık problemi varsa bunu gidermek için belirli yöntemlere başvurulması gerekmektedir. Doğrusallık probleminin temelinde eğer ki doğrusallık eksik ise araştırmacı için analizinde sağladığı bilgilerin yetersiz olduğunu gösterir. Araştımacı doğrusalığın olduğunu anlamak için çok değişkenli normallik varsayımında değişken çiflerinin arasındaki ilişkinin doğrusal olduğunu vurgulaması durumudur. [39]

Bu bilgiler doğrultusunda faktör analizine geçilmeden önce bunlara bakılmalı ve uygun yöntemler ile problemler ortadan kaldırılmalıdır.

2.4 Faktör Analizinin Ortak Entropide Uygulanması

Bu bölümde, temel bileşenler analizi üzerinde birleşik entropi ölçüsünün kullanımını ele almıştır.

Yapılan araştırmalar gösteriyor ki, içinde çok fazla değişkenin yer aldığı durumlarda genellikle çok değişkenli istatistiksel analiz teknikleri tercih edilmektedir. Araştırmacılar bu teknikleri kullanarak, değişkenler arasındaki karmaşık ilişkileri anlamaya çalışmaktadırlar. Analize başlamadan önce ilk yapılması gereken şey değişkenler arasındaki ilişkileri iyi bir şekilde kavrayabilmektir. Sınıflandırılmış verilerin olduğu durumlar için farklı yöntemler

geliştirilmiştir. Bu tür veriler üzerinde çalışılırken, doğrusal olmayan yaklaşımların veya sınıflandırılmış değişkenlerin seviyelerinin kukla değişken olarak kullanılması oldukça yaygın kullanılan yöntemlerdendir. Entropiye bağlı hesaplamalarda ise, değişkenler hakkındaki yapılan hesaplamaların kesin sonuçları yerine sınıflandırılmış verilerin olasılıkları dikkate alınarak işlem yapılmaktadır. Bu gibi durumlarda analizin belirsizlik düzeyi veya analizin düzensizliği hesaplanarak, değişkenler arasındaki ilişki çözümlenmeye çalışılmaktadır. Entropiye bağlı hesaplamaları kullanarak değişkenlerin yapısını anlayabilmek ve verileri daha sade bir biçimde yazabilmek için temel bileşenler analizi kullanılmaktadır.

Analiz yaparken çok sayıda değişkenin kullanılması, araştırmacının işini çok daha karmaşık ve güç bir duruma getirebilmektedir. Bu gibi zorlukların önüne geçmek ve analizi daha kolay yönetilebilir hale getirmek, araştırmacının işini çok daha kolay hale getirecektir. Bu sebeple, çok sayıda değişkenin daha az sayıda değişkene indirgenmeye çalışılması oldukça önemlidir. Daha az değişken olmasıyla birlikte, analizin uygulanabilirliği yüksek ölçüde arttırılabilir ve daha net, daha kolay yorumlanabilir sonuçlar elde edilebilmektedir. Değişken sayısının indirgenmesine yardımcı olan yöntemlerden biri temel bileşenler analizidir. Bu yöntemle değişkenler daha az sayıdaki temel bileşenlerine kadar indirgenmekte ve veri seti özet haline getirilmektedir. [9]

Faktör analizi ile birlikte, p tane değişkeninin bulunduğu durumlarda birbirleriyle ilişkili olan değişkenlerin birbirbirleriyle yanyana getirilmesi ve daha az sayıda elde edilecek olan yeni (ortak) ilişkisi olmayan değişkenlerin ortaya çıkması amaçlanmaktadır. Bu sebeple faktör analizi, boyut indirgeme konusunda, araştırmacılar arasında oldukça popüler olan bir seçenektir. [12]

Faktör analizleri yapılır iken, tüm değişkenler bir arada ele alınmaktadır. Sonrasında ortaya çıkabilecek herhangi bir faktör ise, gözlemlenmiş değişkenlerin bir fonksiyonu olarak yani bağımsız bir değişken olarak kabul edilmektedir. [10]

Entropi ve varyans benzer yapıya sahiptir. Benzer özelliklere sahip olan entropi ve varyansı incelediğimizde ilk olarak ancak ve ancak p_i'den biri 1 ise, o zaman H= O. Bu özellik varyans içinde geçerlidir. Bir diğer ortak özellik ise x ve y değişkenleri bağımsız ise H(x,y) = H(x) + H(y) şeklindedir. Bu durum da varyans

içinde geçerlidir. x ve y bağımsız ise var(xy) = var(x) + var(y) şeklinde ifade edilebilir. [40]

Belirsizlik ölçüsü için varyans ve entropi arasındaki ilişki ve farklılıklar mevcuttur. Bilgi teorisi çalışmaları ve uygulamaları sırasında, farklı verilere göre varyans ve entropi arasında uygun bir istatistik seçilmelidir. Bu aşamada Tablo 2.2'den faydalanılabilmektedir. [40]

Tablo 2.2 Veri setinde kullanılan değişkenler

Veri Tipi	**İstatistik**
Değerler (frekans dağılımında dönüştürülmesi güç değerler)	Varyans
Değerler ve olasılıklar	Varyans ya da entropi
Olasılıklar (kukla değişken uygulamak kolay)	Varyans ya da entropi
Olasılıklar (kukla değişken uygulamak zor)	Entropi

Ortak entropi hesaplanmasında kullanılan formül (2.13) ile gösterilmiştir.

$$H(X,Y) = -\sum_{i,j} p(i,j)\, log\, p(i,j) \tag{2.13}$$

Bu formül önceki bölümde gösterilmişti. Bu aşamada bir X_α rassal değişkeni ele alındığında H(X_α, X_α) değerinin hesaplanması istenilirse H(X_α, X_α) = $-\sum_{i,j} p(i,i)\, log\, p(i,i)$ biçiminde yazılacaktır. $p(i,i)$ değeri ise $p(i)$ değerine eşit olduğundan dolayı (2.14) ile gösterildiği şekilde hesaplanmaktadır.

$$H(X,X) = -\sum_{i} p(i)\, log\, p(i) \tag{2.14}$$

Buradaki sonuç ise X_α rassal değişkeninin marjinal entropisine eşittir. Bu durumda ulaşılacak sonuç ise

$$H(X,X) = H(X) \tag{2.15}$$

(2.15) eşitliğidir. [31]

Birleşik entropi ile her bir olayın ayrı entropileri arasındaki ilişkiyi incelediğimizde ise (2.16) ile ifade edilir.

$$H(X,Y) \leq H(X) + H(Y) \tag{2.16}$$

Bu eşitsizlik Schwarz-Cauchy eşitsizliğine denktir. Bu eşitsizliğe ulaşabilmemizdeki sebep olayın bağımsız olması sebebiyle p(i,j)=p(i)p(j) denkleminin geçerli olmasıdır. Sonuç olarak birleşik entropi olayının belirsizliği, X ve Y'nin belirsizliklerinin ayrı ayrı toplamına eşit veya daha küçüktür. [31]

Bir veri setinde bağımlılık ölçüsünün hesaplanmasında ise (2.16) eşitliğinden faydalanarak;

$$\beta_{x_\alpha,x_\beta} \leq \frac{H(x_\alpha) + H(x_\beta)}{H(x_\alpha, x_\beta)} \tag{2.17}$$

(2.17) ile bağımlılık ölçüsünün hesaplanması mümkündür. [21]

Benzer bir çıkarım, entropi tabanlı "karşılıklı bilgi" fikrini kullanarak elde edilebilir. Bu, iki tesadüfi değişken arasındaki bağımlılığın bir ölçüsünü sağlar. Karşılıklı bilgi tipik olarak (2.18) ile temsil edilir.

$$I(x_\alpha, x_\beta) = H(x_\alpha) + H(x_\beta) - H(x_\alpha, x_\beta) \tag{2.18}$$

Karşılıklı bilgi formülü aynı zamanda birleşik entropi matrisinin standartlaştırılması işlemi ile ilişkilendirilebilir. Bunun yanı sıra, x_α ve x_β değişkenleri arasındaki entropi-korelasyon katsayıları da hesaplanabilir. [41]

Değişkenlerin entropi değerleri ile birleşik entropi değerleri elde edildikten sonra, p × p boyutlarında simetrik bir birleşik entropi matrisi oluşturmak mümkündür. [42]

	X_1	...	X_p
X_1	$H(X_1)^*$	...	$H(X_1, X_p)$
⋮	⋮	⋱	⋮
X_p	$H(X_1, X_p)$	...	$H(X_p)$

Şekil 2.21 Birleşik entropi matrisinin görselleşitrilmiş hali

Oluşturulan birleşik entropi matrisi Şekil 2.21'deki gibi görüneccektir. Bu noktadan itibaren, birleşik entropi matrisi elde edildikten sonra, standardize edilmiş matrisin özdeğerleri belirlenerek benzer bir faktör analizi süreci uygulanacaktır.

3
UYGULAMA

3.1 Amaç

Bu çalışma doğrultusunda ülkelerin sosyo-ekonomik ve sağlık alanındaki faktörleşmesi üzerinde durulması amaçlanmıştır. Literatürde faktör analizi ve entropi alanında çeşitli araştırmalar mevcuttur. Ancak istatistiksel anlamda yapılan entropi tabanlı analizi yöntemlerinin uygulamaları açısından lüteratür malesef oldukça zengin değildir. Bu sebeple bu tez çalışmasıyla birlikte istatistiksel anlamda entropi tabanlı bir uygulama ile hem dikkat çekilmesi hem de literatüre katkıda bulunması amaçlanmaktadır.

Bu çalışmada hedef, veri setini analiz etmek ve veri setindeki temel faktörleri belirlemek için faktör analizi ve entropi tabanlı bir yaklaşım olan birleşik entropi yönteminin kullanılmasıdır. Bu uygulama kapsamında, her iki yöntem de kullanılarak elde edilen faktörlerin belirlenmesi hedeflenmiştir. Bu bağlamda, her iki analiz yöntemiyle elde edilen faktörler karşılaştırılması amaçlanmıştır. Bu karşılaştırma, farklı analiz yöntemlerinin veri setindeki yapıyı anlama ve yorumlama yeteneklerini değerlendirmeyi amaçlamıştır.

Sonuç olarak, bu çalışma ile hem faktör analizi hem de entropi tabanlı bir yaklaşım olan birleşik entropi kullanarak veri setindeki yapıyı incelemiş ve bu iki farklı analiz yönteminin sonuçlarını karşılaştırarak araştırmanın genel amacını oluşturmuştur.

3.2 Veri Seti

Bu çalışma Kaggle üzerinden “Unsupervised Learning on Country Data” başlıklı veri seti alınarak yapılmıştır. Veri setinin ülkeler bazında genel gelişimini belirleyen sosyo-ekonomik ve sağlık faktörtörlerini kullanarak katagorize edilmesi amaçlanmıştır. Bu bağlamda 167 ülke üzerinden bir değerlendirme söz konusudur. Veri setinde bulunan değişkenler sırasıyla çocuk ölüm oranı, ihracat, sağlık, ithalat, gelir, enflasyon, ortalama yaşam süresi, doğurganlık oranı ve kişi başına gayri safi yurtiçi hasıla(gdpp) şeklinde 9 değişkendir. Bu bağlamda Tablo 3.1’de görüldüğü

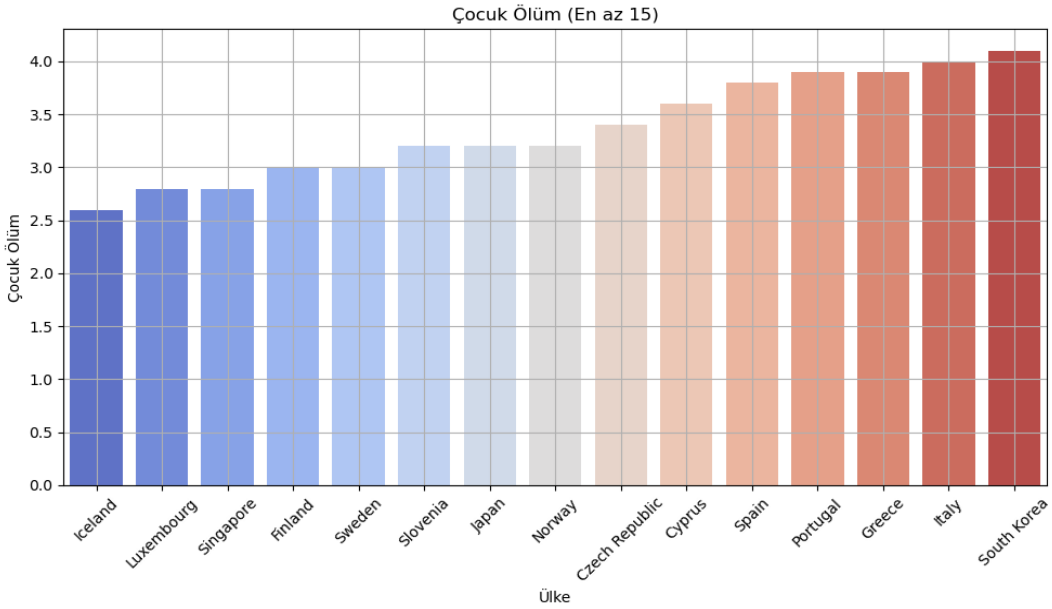
gibi hem değişkenler hem de birleşik entropi yöntemi yapılırken kullanılan değişken isimleri yer almaktadır.

Tablo 3.1 Veri setinde kullanılan değişkenler

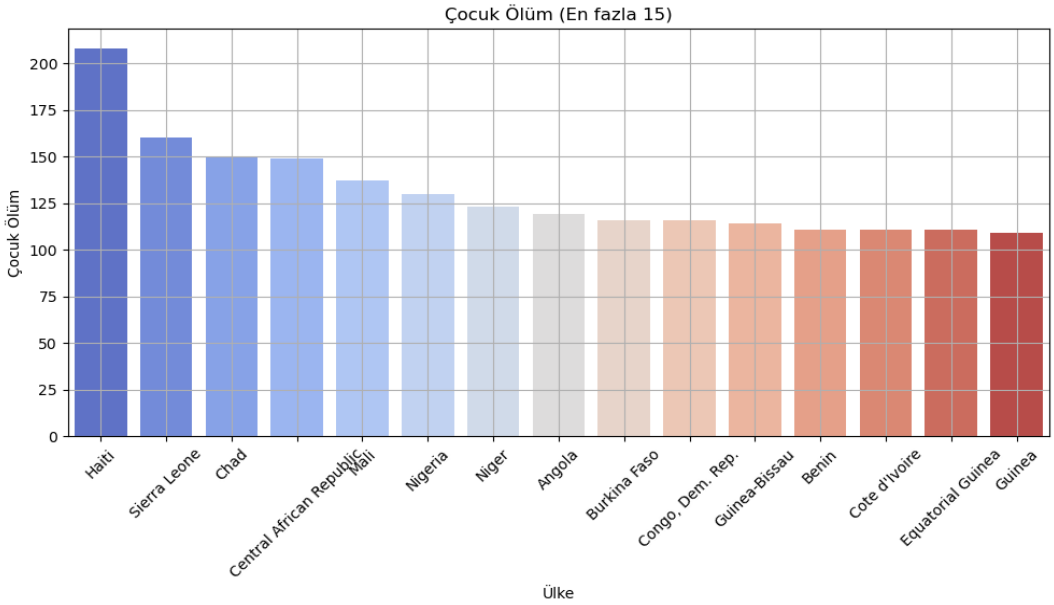
Değişkenler	**Değişken Adı**	**Açıklama**
Çocuk Ölümü	X1	Her 1000 doğumda, 5 yaş altı ölen çocuk ölümleri.
İhracat	X2	Kişi başına düşen mal ve hizmet ihracatı. Kişi başına düşen GSYİH'nin yüzdesi olarak verilmiştir.
Sağlık	X3	Kişi başına düşen toplam sağlık harcamaları. GSYİH'nin yüzdesi olarak verilmiştir.
İthalat	X4	Kişi başına düşen mal ve hizmet ithalatı. Kişi başına düşen GSYİH'nin yüzdesi olarak verilmiştir.
Gelir	X5	Kişi başına net gelir.
Enflasyon	X6	Toplam GSYİH'nin yıllık büyüme oranı
Beklenen Yaşam Süresi	X7	Mevcut ölüm oranları aynı kalsa, yeni doğan bir çocuğun ortalama yaşayacağı yıl sayısı
Toplam Doğurganlık	X8	Mevcut yaş-doğurganlık oranları aynı kalsa, her kadına doğacak çocuk sayısı
gdpp	X9	Kişi başına düşen GSYİH. Toplam GSYİH'nin toplam nüfusa bölünmesiyle hesaplanır.

Tablo 3.1'deki değişkenler ele alındığında ülkelerin sosyo-ekonomik ve sağlık açısından değerlendirilmesi amacıyla ilk olarak grafiksel yorumlara dayalı inceleme yapılmıştır. Bu bağlamda her değişken için ayrı ayrı ilk 15 ve son 15 ülke olacak şekilde incelenmiştir. Bu inceleme sırasıyla çocuk ölümü, ihracat, sağlık, ithalat, gelir, enflasyon, beklenen yaşam süresi, toplam doğurganlık, kişi başına düşen gayri safi yurtiçi hasılat şeklinde olacaktır.

İlk olarak ülkelerde gözlemlenen çocuk ölüm oranı incelenmiştir. Söz konusu değişkenin verisinde bahsedilen oran ise her 1000 çocuktan 5 yaşından küçük olanların ölüm oranıdır. İlk olarak bu oranın en az olduğu ilk 15 ülke akabinde ise bu oranın en yüksek olduğu 15 ülke incelenmiştir.



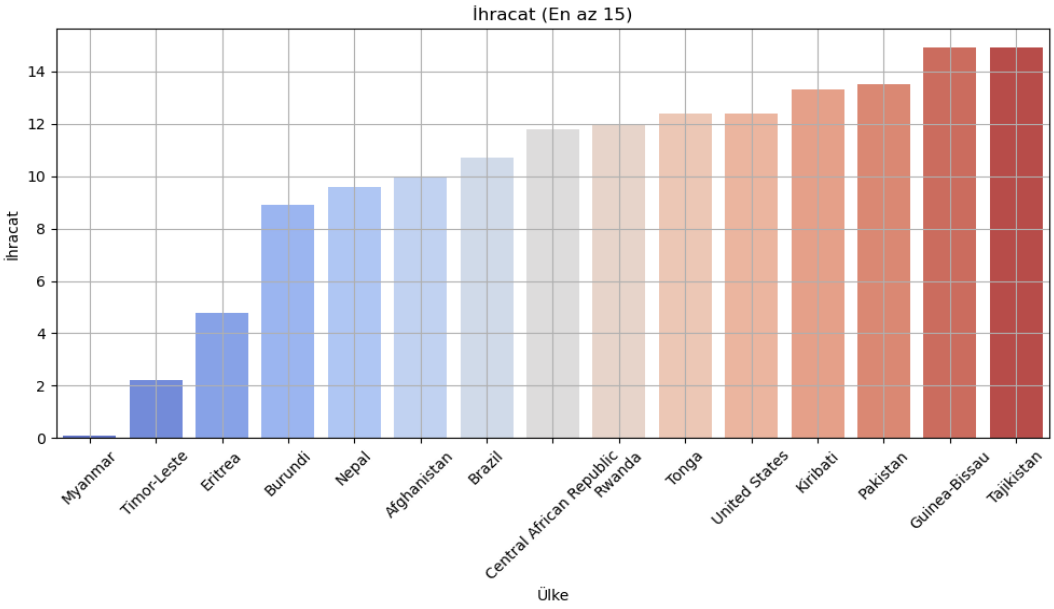
Şekil 3.1 Çocuk ölümünün en az olduğu 15 ülke



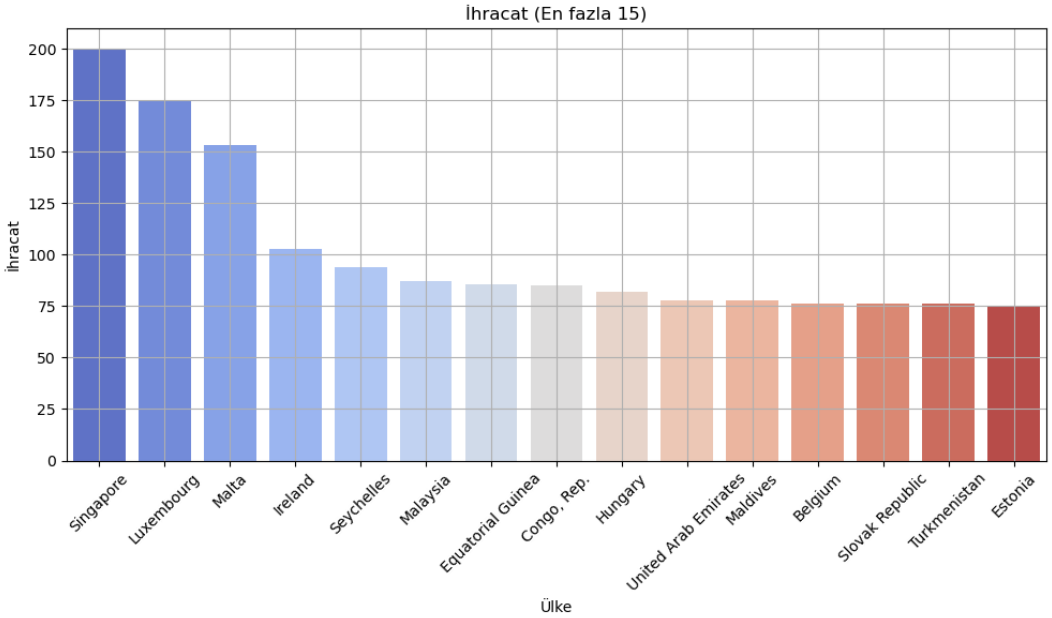
Şekil 3.2 Çocuk ölümünün en çok olduğu 15 ülke

Bu grafikler doğrultusunda çocuk ölümünün en çok olduğu ülke her 1000 canlıdan 200 üstünde 5 yaşının altında olan çocuklarının ölümü olduğu görülürken, İzlanda'da ise her 1000 canlıdan 5 yaşından küçük olan sadece 3 çocuk ölümü gözlemlenmektedir.

İhracat değişkeni incelendiğinde ise bu bağlamda ele alınan konu kişi başına düşen mal ve hizmet ihracatı. İlgili değişken GSYİH'nin yüzdesi olarak verilmektedir.

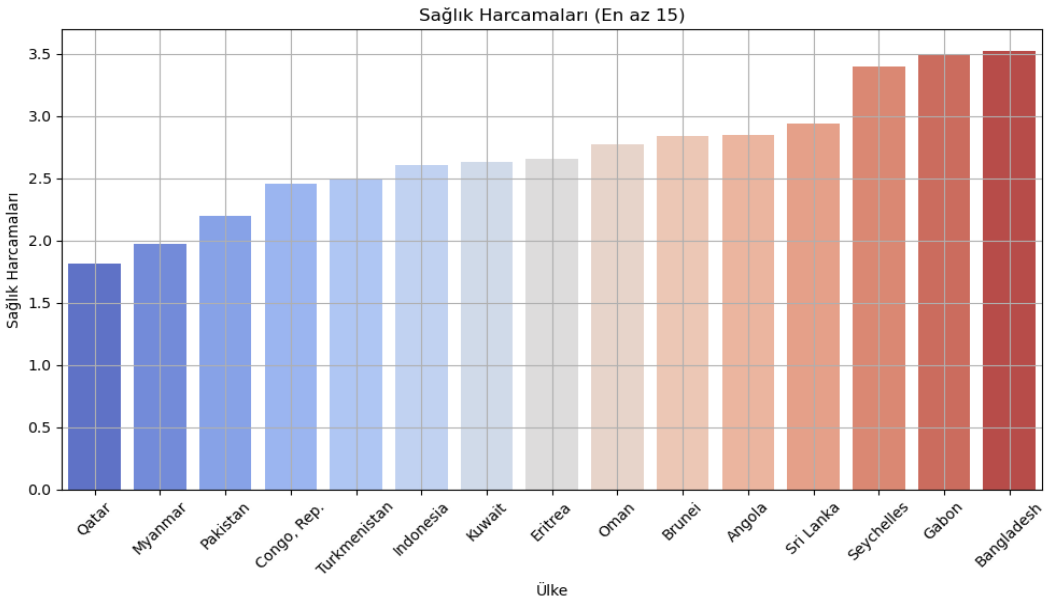


Şekil 3.3 İhracat oranının en az olduğu 15 ülke

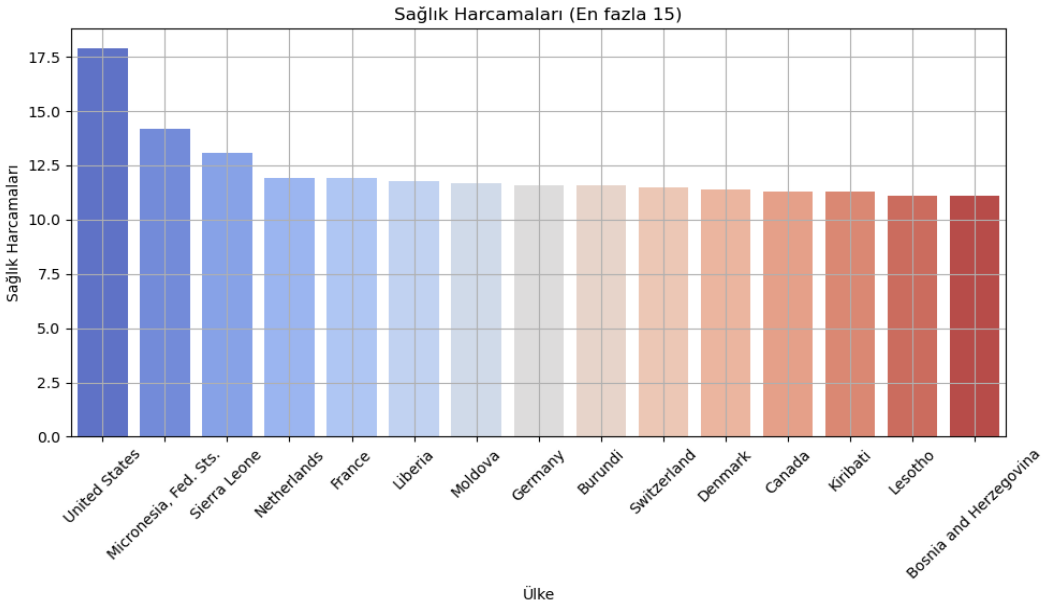


Şekil 3.4 İhracat oranının en çok olduğu 15 ülke

Sağlık değişkeni incelendiğinde ise bu bağlamda ele alınan konu Kişi başına düşen toplam sağlık harcaması ele alınmıştır. İlgili değişken GSYİH'nin yüzdesi olarak verilmektedir. İlgili değişken incelendiğinde ise Şekil 3.5 ve Şekil 3.6'da görüldüğü gibidir.

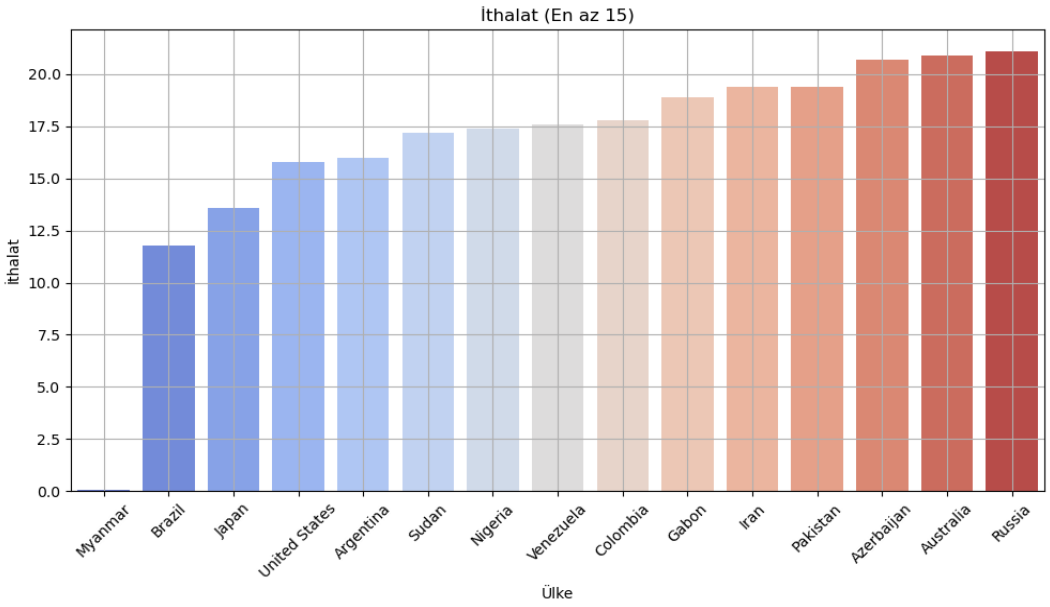


Şekil 3.5 Sağlık harcamaları oranının en az olduğu 15 ülke

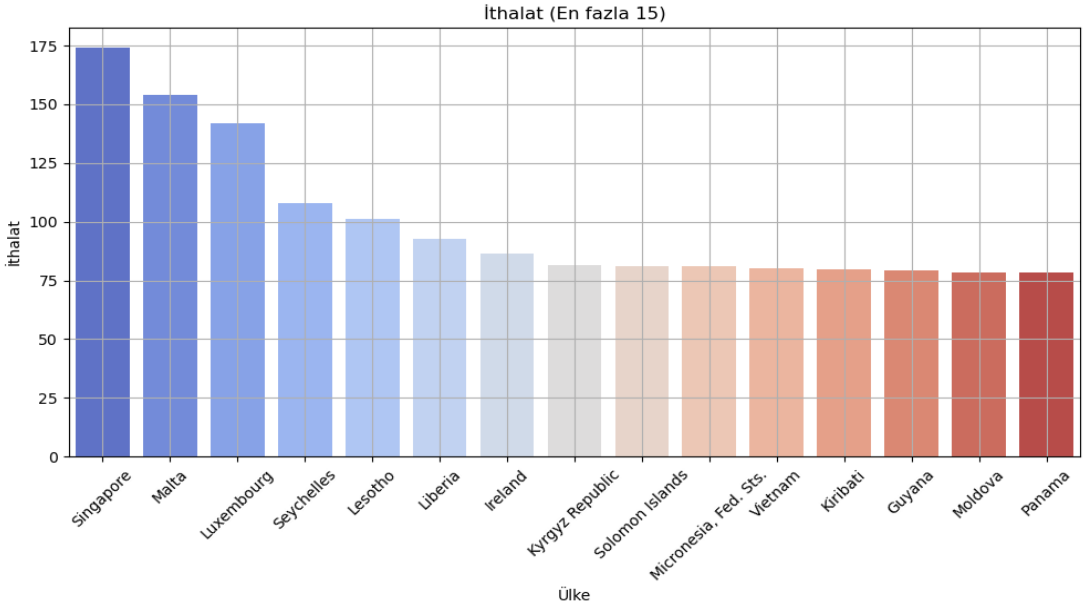


Şekil 3.6 Sağlık harcamaları oranının en çok olduğu 15 ülke

Bir sonraki değişken olan ithalat değişkeni kişi başına düşen mal ve hizmet ithalatı ele alınmıştır. İlgili değişken GSYİH'nin yüzdesi olarak verilmektedir.

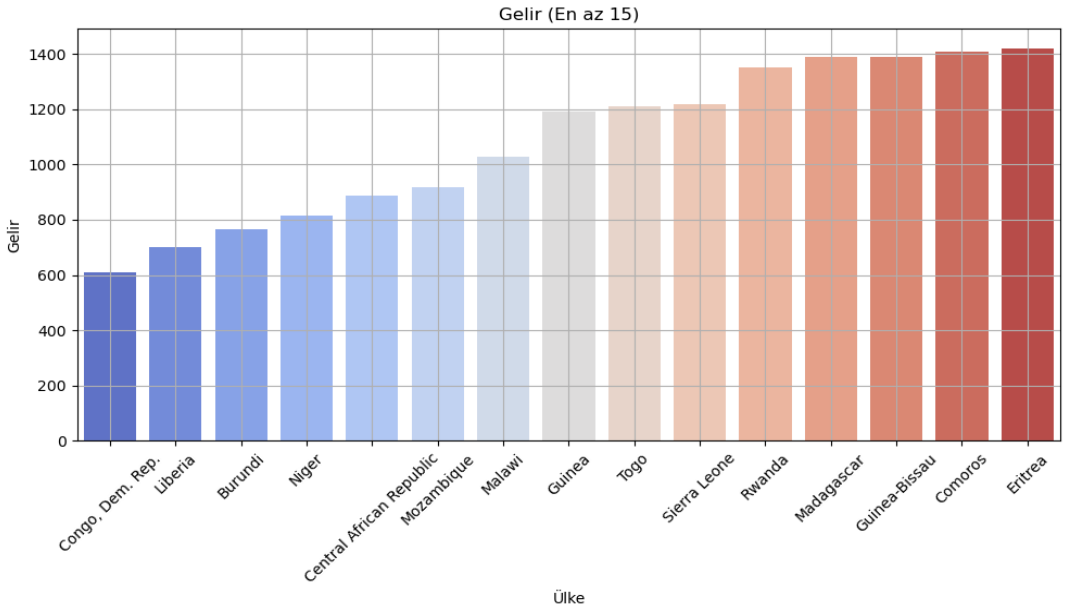


Şekil 3.7 İthalat oranının en az olduğu 15 ülke

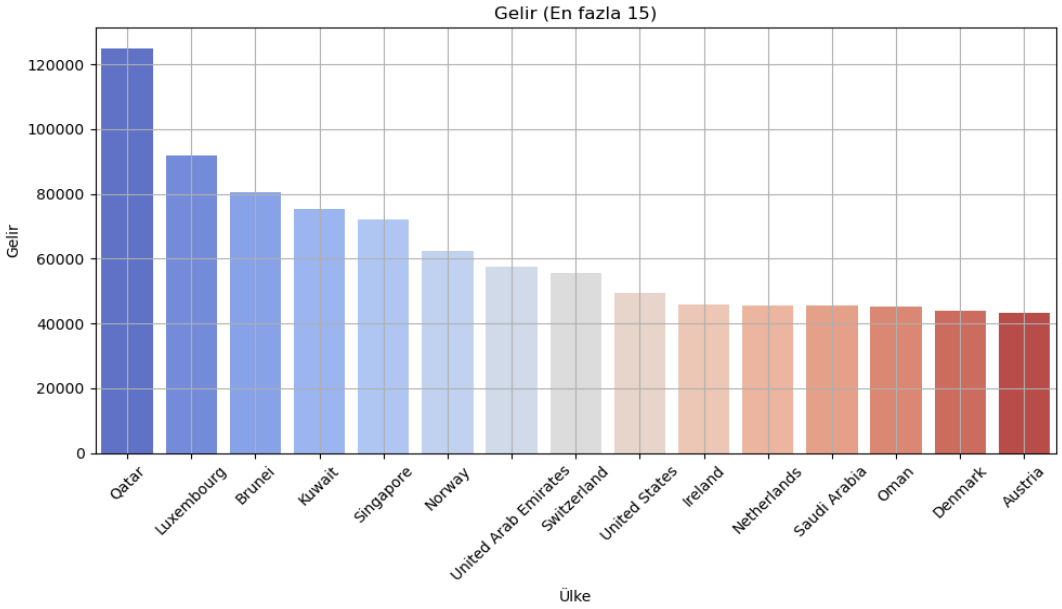


Şekil 3.8 İthalat oranının en çok olduğu 15 ülke

Buna takiben bir sonraki değişken olan gelir değişkeni ise kişi başına net gelir olarak incelenmiştir. Bu bağlamda geliri en çok olan ülke Katar iken en düşük gelire sahip ülke Kongo Demokratik Cumhuriyeti olarak gözlemlenmiştir.

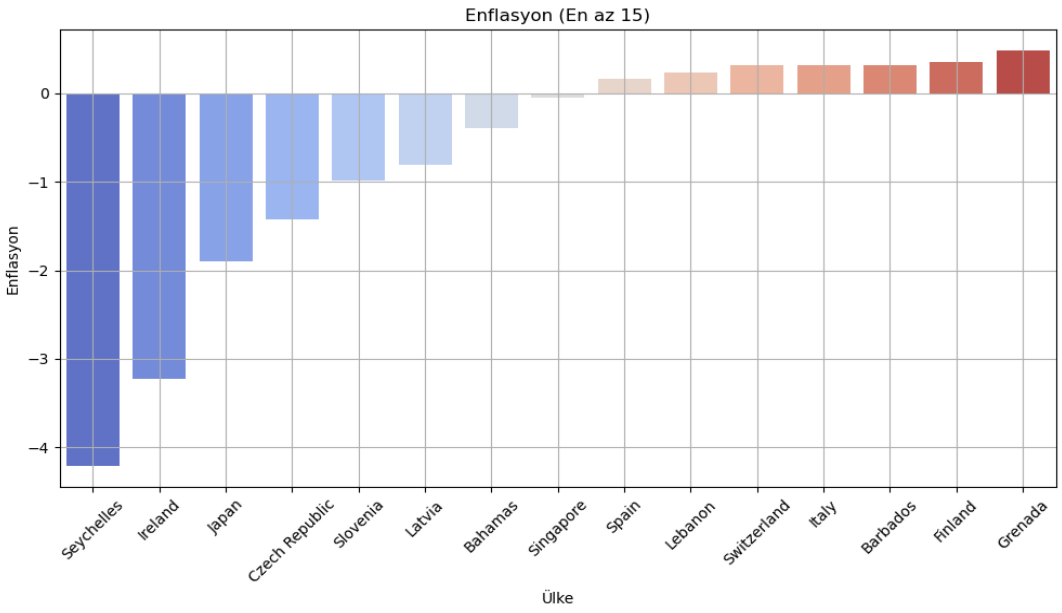


Şekil 3.9 Gelirin en az olduğu 15 ülke

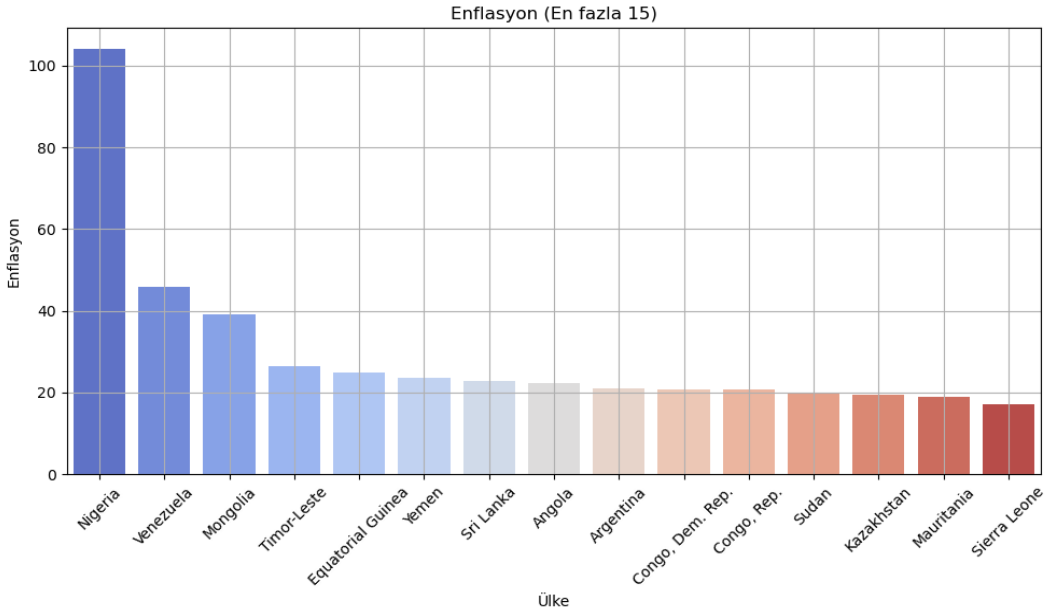


Şekil 3.10 Gelirin en çok olduğu 15 ülke

Enflasyon değişkeni ele alındığında enflasyonun en yüksek olduğu ülkelerin başında Nijerya, Venezuela ve Moğolistan iken Şeyşeyler, İrlanda ve Japonda ise en düşük seviyede olduğu Şekil 3.11 ve Şekil 3.12'de incelenmiştir. Bu değişekne toplam GSYİH'nin yıllık büyüme oranı ölçümünü göstermektedir.

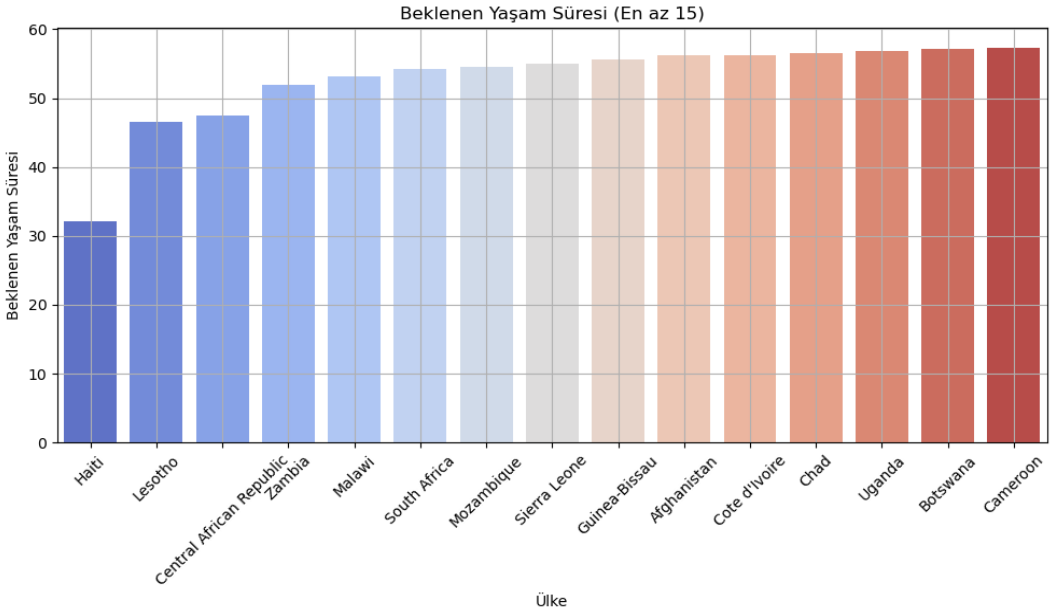


Şekil 3.11 Enflasyonun en düşük olduğu 15 ülke

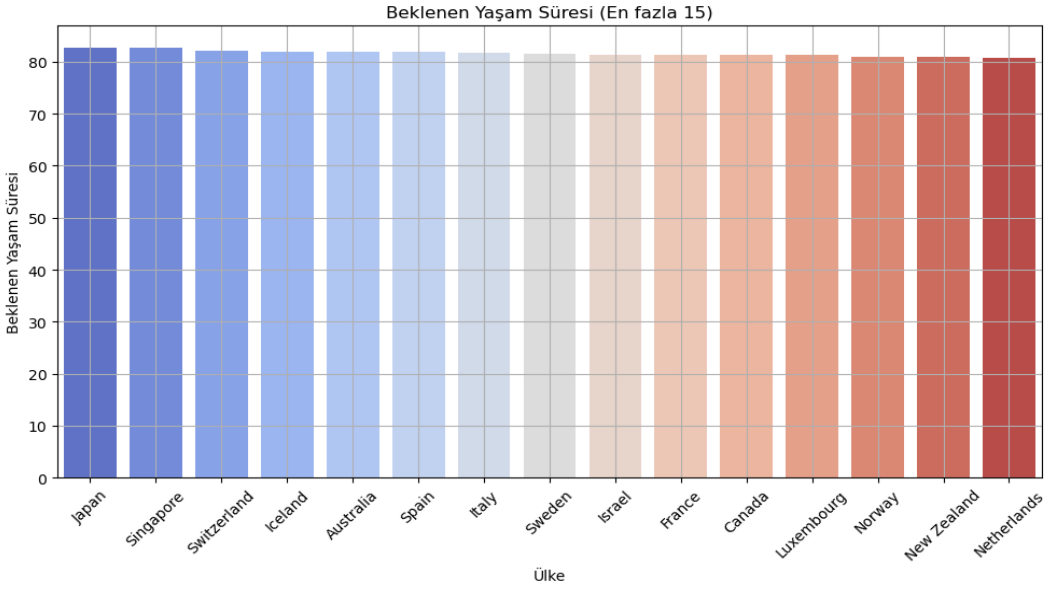


Şekil 3.12 Enflasyonun en yüksek olduğu 15 ülke

Bir sonraki değişken olan yaşam beklentisi değişkeninde ise eğer mevcut ölüm oranı kalırsa, yeni doğan bir çocuğun ortalama yaşayacağı yıl sayısı olarak hesaplanmıştır. Bu bağlamda bahsi geçen değişken Şekil 3.13 ve Şekil 3.14'te incelenmiştir. Beklenen yaşam sürelerinin incelendiği bu değişkende beklenen en kısa yaşam süresi Haiti iken, beklenen en uzun yaşam süresi Japonya olmuştur.

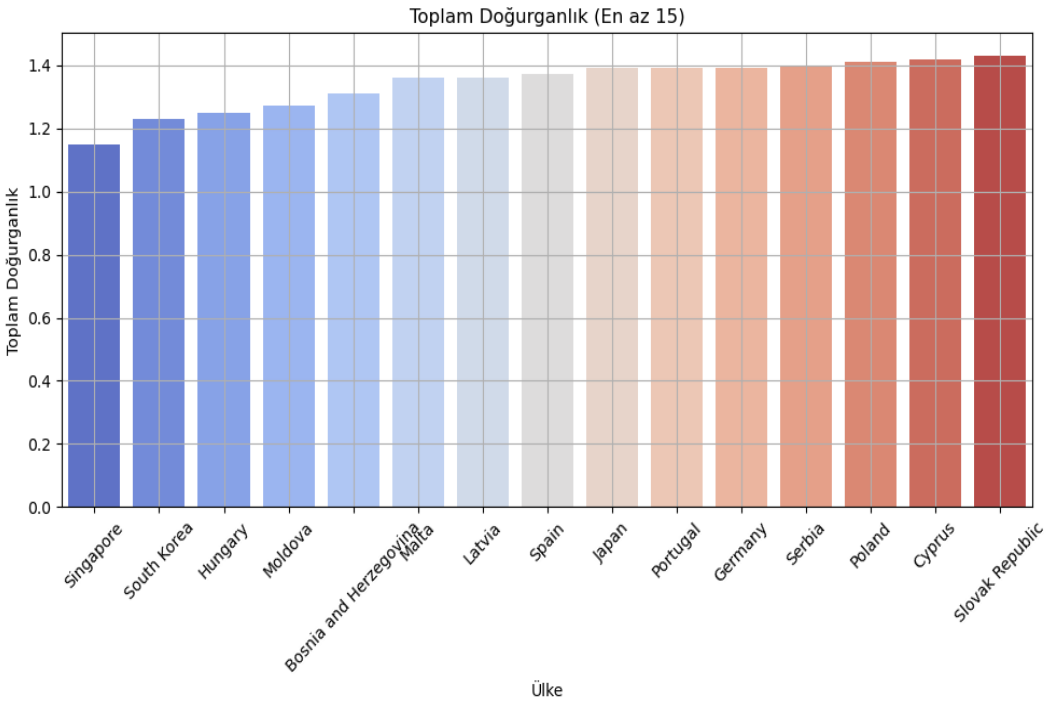


Şekil 3.13 Beklenen yaşam süresi en düşük olduğu 15 ülke

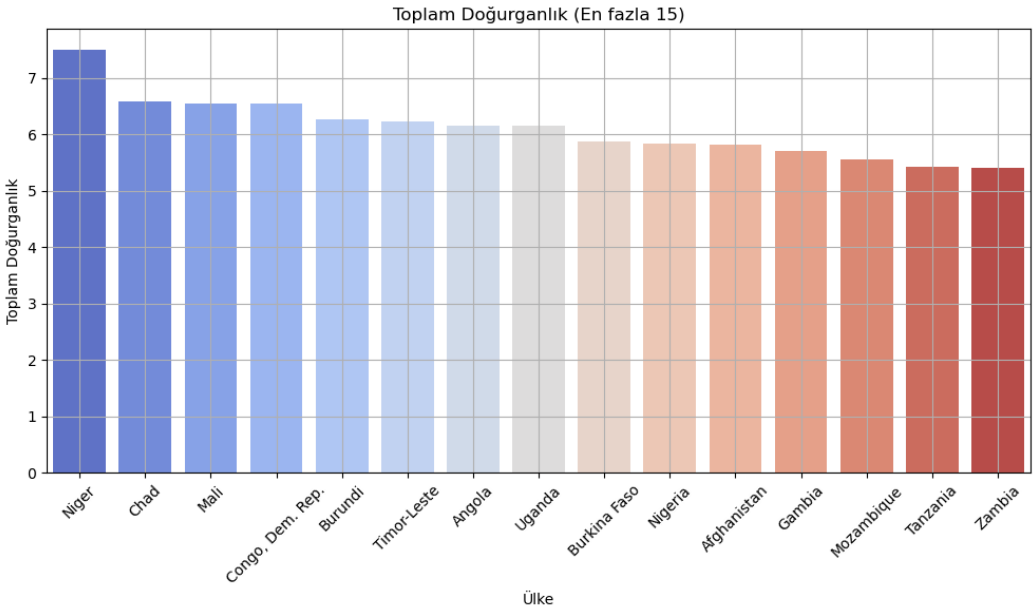


Şekil 3.14 Beklenen yaşam süresi en yüksek olduğu 15 ülke

Toplam doğurganlık değişkeni ele alındığında Şekil 3.15 ve Şekil 3.16'da incelenmiştir. Bahsi geçen değişken ile mevcut yaş-doğurganlık oranları aynı kalırsa, her kadının doğuracağı çocuk sayısı ele alınmıştır. Bu bağlamda Singapur, Güney Kore, Macaristan gibi ülkelerde doğurganlık düşük iken, Nijer, Çad, Mali gibi ülkelerde doğurganlık yüksek seviyede olduğu gözlemlenmiştir.

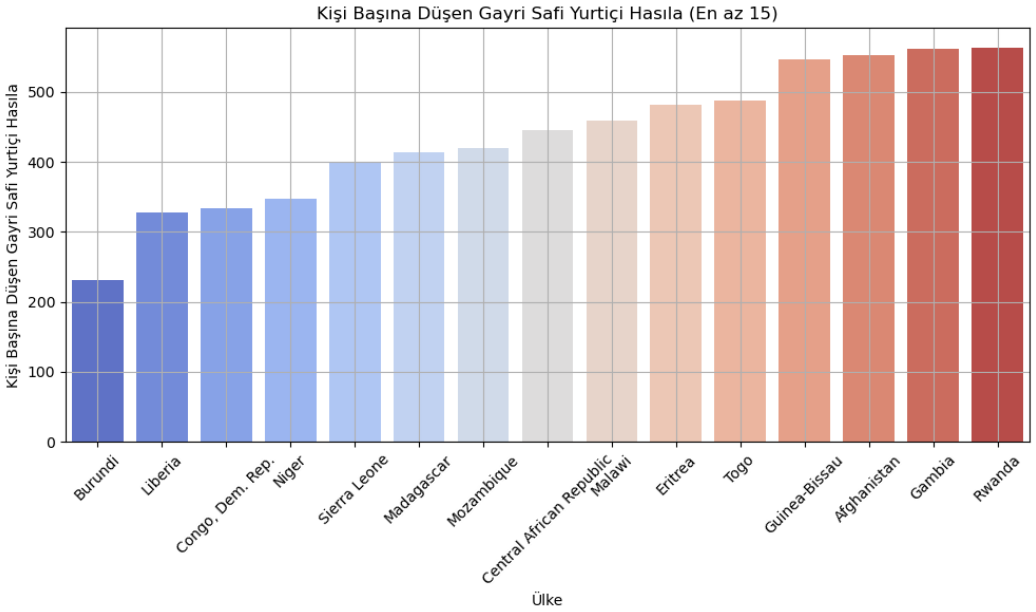


Şekil 3.15 Doğruganlık düzeyinin en düşük olduğu 15 ülke

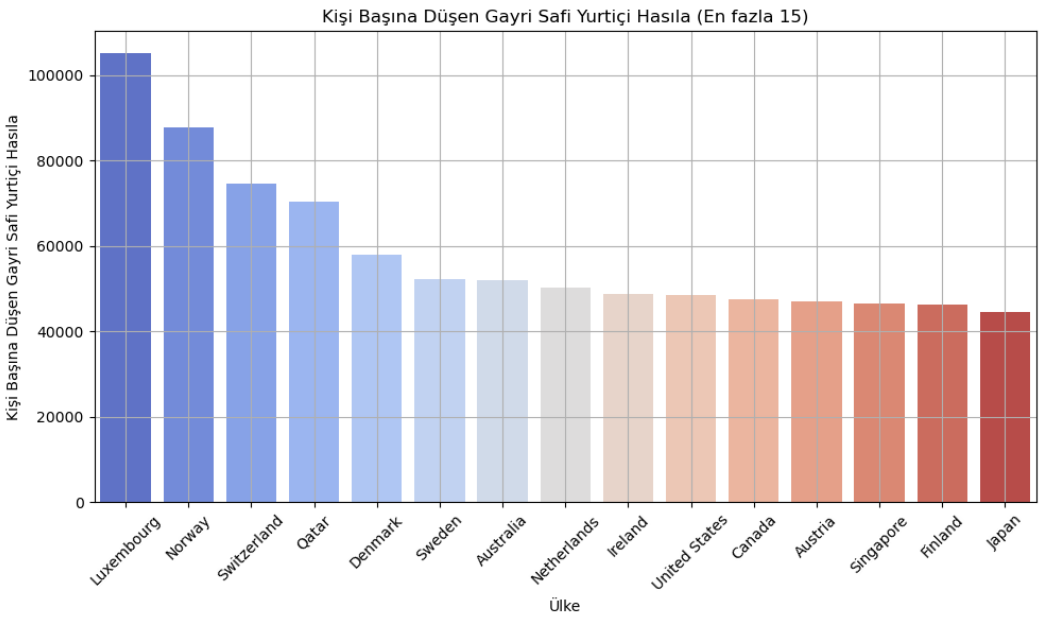


Şekil 3.16 Doğurganlık düzeyinin en yüksek olduğu 15 ülke

Kişi başına düşen gayri safi yurtiçi hasıla değişkeni ele alındığında ise bu değişken toplam GSYİH'nin toplam nüfusa bölünmesiyle hesaplanmıştır. Bu kapsamda Şekil 3.17 ve Şekil 3.18'de gösterilmiştir.



Şekil 3.17 Kişi başına düşen gayri safi yurtiçi hasılat düzeyinin en düşük olduğu 15 ülke



Şekil 3.18 Kişi başına düşen gayri safi yurtiçi hasılat düzeyinin en yüksek olduğu 15 ülke

Yapılan bu grafiksel incelemelerin bir sonraki başlıklarda yapılacak olan ortak entropi ve faktör analizi uygulamasında elde edilecek olan faktörleşmeler sonucunda faktörlerin isimlendirilmesinde fayda sağlayacaktır.

3.3 Faktör Analizini Uygulaması

Bu başlık altında ilk olarak veri seti SPSS ile faktör analizine sokulup kaç faktöre indirgendiği hesaplanmıştır ve isimlendirilmiştir. İlk aşamada SPSS ile faktör analizi incelemesi yapılmıştır. Bu bağlamda ilk olarak KMO değeri hesaplanmış olup, bu değer 0.68 ile orta seviyede çıkmıştır. Tablo 3.2'de görüldüğü gibi KMO değerinin faktör analizi için uygun olduğunu göstermektedir.

Tablo 3.2 Kaiser-Meyer-Olkin ve Bartlett Testi

Kaiser-Meyer-Olkin ve Bartlett Testi

Kaiser-Meyer-Olkin Örnekleme Yeterliliğinin Ölçüsü.		0,678
Bartlett'in Küresellik Testi	Tahmini Ki-Kare	1169,737
	Serbestlik Derecesi	36
	p-değeri	0,000

Sonrasında ise değişkenler incelenmiş olup bu kapsamda veri setindeki 9 değişkenin de analiz sonucunda kalması kararına varılmıştır. Buna Tablo 3.3'te görüldüğü gibi 0.5 altında değer olmaması sebebiyle karar verilmiştir.

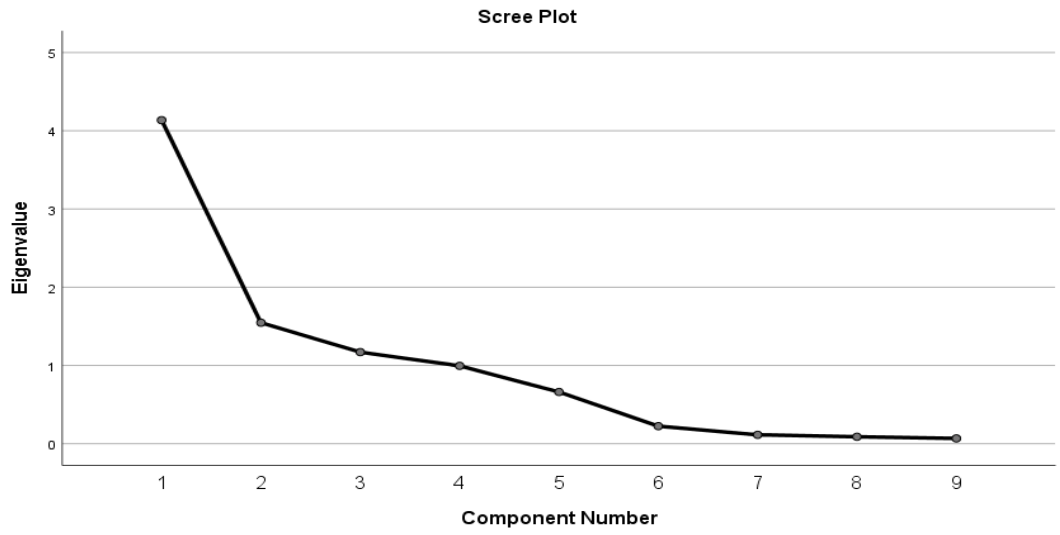
Tablo 3.3 Toplam Varyans

Toplamsal varyans

	Initial	Extraction
cocuk_olumu	1.000	.786
ihracat	1.000	.939
saglik	1.000	.602
ithalat	1.000	.911
gelir	1.000	.764
enflasyon	1.000	.638
yasam_suresi	1.000	.842
toplam_dogurganlik	1.000	.712
gdpp	1.000	.659

Extraction Method: Principal Component Analysis.

Faktör analizine geçiş yapılmasının ardından ise Şekil 3.19'da gösterilmiş olan Scree Plot ile 3 faktöre inidirgenmesine gerektiği gözlemlenmektedir. Scree Plot'a bakıldığında faktör sayısının 3 olarak belirlenmesinin sebebi özdeğerin 1'den büyük olan değerler kadar faktöre indirgeneceğidir. Bu kapsamda özdeğerin 1'den büyük olduğunu gösteren 3 nokta mevcuttur. Bu bilgilere ek olarak indirgenen faktörlerin açılanan toplam varyansı ise Şekil 3.20 ile gösterilmektedir.



Şekil 3.19 Scree plot

Açıklanan Toplam Varyans

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.136	45.952	45.952	3.666	40.734	40.734
2	1.546	17.182	63.133	1.794	19.938	60.672
3	1.170	13.004	76.138	1.392	15.466	76.138
4	.995	11.053	87.191			
5	.661	7.340	94.531			
6	.224	2.484	97.015			
7	.113	1.260	98.276			
8	.088	.981	99.257			
9	.067	.743	100.000			

Extraction Method: Principal Component Analysis.

Şekil 3.20 Açıklanan toplam varyans

Şekil 3.19'da ve Şekil 3.20'de görüldüğü gibi faktör analizi ile 9 değişkenden 3 faktöre indirgenmiştir. Bu bağlamda ilk faktörün açıklanan toplam varyansı %40.7 iken ikincinin %19.9 olduğu gözükmektedir. Bu bağlamda ilk iki faktörün açıklanabilir toplam varyansının %60.6 olduğu gözlemlenmektedir. Üçüncü faktörün ise %15.5 iken üç faktörün toplam açıklanabilir varyans oranının %76.1 olduğu gözlemlenmektedir. Faktör analizinde Varimax dönüşümü kullanılmıştır.

Tablo 3.4 SPSS ile faktör analizinden elde edilen faktörler

	Sosyal	**Ekonomi**	**Sağlık**
yasam_suresi	0.902	-0.005	-0.169
cocuk_olumu	-0.852	-0.037	0.244
gelir	0.834	0.241	0.100
toplam_dogurganlik	-0.793	-0.078	0.276
gdpp	0.785	0.180	-0.097
ithalat	-0.022	0.929	-0.219
ihracat	0.375	0.880	0.154
enflasyon	-0.138	-0.216	0.756
saglik	0.158	-0.111	-0.752

Mevcut 9 adet değişkenin 3 faktör ile temsil edilebileceğine karar verilir. Bu 3 faktör Tablo 3.4'te gösterilmiştir. Bu faktörlerin isimlendirilmesi ise değişkenlere bağlı olarak yapılmıştır. Bu bağlamda faktör 1'i temsilen sosyal faktörü olarak

isimlendirilmiştir. Sosyal faktöründeki değişkenler ise sırasıyla yaşam süresi, çocuk ölümü, gelir, toplam doğurganlık ve gdpp (kişi başına düşen gayri safi yurtiçi hasılat) süresi şeklindedir. Faktör 2'yi temsilen ekonomi faktörü ele alınmıştır. Bahsi geçen faktördeki değişkenler ise ithalat ve ihracat değişkenleridir. Son olarak faktör 3'ü temsilen ise sağlık faktöründe sadece sağlık değişkeni bulunmaktadır.

3.4 Ortak Entropi Uygulaması

Bu başlık altında ilk olarak veri seti Python ve Excel yardımları ile kaç faktöre indirgendiği hesaplanmıştır ve isimlendirilmiştir. Bu bağlamda ilk olarak veri setine ait olan 9 değişkenin tek tek entropisi hesaplanmaktadır. Bu değerler sırasıyla Tablo 3.5'te gösterilmiştir.

Tablo 3. 5 Değişkenlere ait entropi değerleri

Değişken	**Entropi Değeri**
X1	7.0183
X2	7.1351
X3	7.1351
X4	7.1921
X5	7.2520
X6	7.2474
X7	6.8775
X8	6.9944
X9	7.2520

Tablo 3.6'da ise değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında kararlar gösterilmektedir.

Tablo 3. 6 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar

	$H(X_\alpha, X_\beta)$	$\frac{H(X_\alpha) + H(X_\beta)}{H(X_\alpha, X_\beta)}$	Değişkenler Hakkında Karar
X1,X2	7.952450	1.7798	Bağımlı.
X1,X3	7.998094	1.7696	Bağımlı.

Tablo 3. 7 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar(devamı)

X1,X4	7.954710	1.7864	Bağımlı.
X1,X5	8.135143	1.7542	Bağımlı.
X1,X6	8.029995	1.7766	Bağımlı.
X1,X7	7.890984	1.7610	Bağımlı
X1,X8	7.968946	1.7584	Bağımlı
X1,X9	8.135143	1.7542	Bağımlı
X2,X3	8.117179	1.7580	Bağımlı.
X2,X4	7.964127	1.7990	Bağımlı.
X2,X5	8.193555	1.7559	Bağımlı.
X2,X6	8.117179	1.7719	Bağımlı.
X2,X7	7.949890	1.7626	Bağımlı
X2,X8	8.056506	1.7538	Bağımlı
X2,X9	8.193555	1.7559	Bağımlı
X3,X4	8.157627	1.7563	Bağımlı.
X3,X5	8.193555	1.7559	Bağımlı.
X3,X6	8.092434	1.7773	Bağımlı.
X3,X7	8.006342	1.7502	Bağımlı
X3,X8	7.978662	1.7709	Bağımlı
X3,X9	8.193555	1.7559	Bağımlı
X4,X5	8.222028	1.7568	Bağımlı.
X4,X6	8.169603	1.7675	Bağımlı.
X4,X7	7.889517	1.7833	Bağımlı
X4,X8	8.093227	1.7529	Bağımlı
X4,X9	8.222028	1.7568	Bağımlı
X5,X6	8.249708	1.7576	Bağımlı.
X5,X7	8.064754	1.7520	Bağımlı
X5,X8	8.123167	1.7538	Bağımlı
X5,X9	8.151639	1.7793	Bağımlı
X6,X7	8.062494	1.7519	Bağımlı
X6,X8	7.995534	1.7812	Bağımlı

Tablo 3. 8 Değişken çiftleri için hesaplanan birleşik entropi değerleri ve değişken hakkında karar(devamı)

X6,X9	8.249708	1.7576	Bağımlı
X7,X8	7.935953	1.7480	Bağımlı
X7,X9	8.064754	1.7520	Bağımlı
X8,X9	8.123167	1.7538	Bağımlı

Tablo 3.6'da görüldüğü gibi ortak entropi değerleri hesaplanmıştır. Bu bilgiler ışığında bağımlı olup olmadığı da kontrol edilmiştir. Birleşik entropi değerlerinin 1 ile 2 arasında olması beklenmektedir. [31] Bir sonraki adım ise birleşik entropi matrisinin elde edilmesidir. Birleşik entropi matrisi Tablo 3.7'de gösterilmektedir.

Tablo 3. 7 Birleşik Entropi Matrisi

	1	2	3	4	5	6	7	8	9
1	7.01832	7.95245	7.99809	7.95471	8.13514	8.03000	7.89098	7.96895	8.13514
2	7.95245	7.13514	8.11718	7.96413	8.19356	8.11718	7.94989	8.05651	8.19356
3	7.99809	8.11718	7.13514	8.15763	8.19356	8.09243	8.00634	7.97866	8.19356
4	7.95471	7.96413	8.15763	7.19209	8.22203	8.16960	7.88952	8.09323	8.22203
5	8.13514	8.19356	8.19356	8.22203	7.25197	8.24971	8.06475	8.12317	8.15164
6	8.03000	8.11718	8.09243	8.16960	8.24971	7.24745	8.06249	7.99553	8.24971
7	7.89098	7.94989	8.00634	7.88952	8.06475	8.06249	6.87754	7.93595	8.06475
8	7.96895	8.05651	7.97866	8.09323	8.12317	7.99553	7.93595	6.99437	8.12317
9	8.13514	8.19356	8.19356	8.22203	8.15164	8.24971	8.06475	8.12317	7.25197

Birleşik entropi matrisi yardımıyla entropi-korelasyon matrisine geçiş yapılcaktır. Bu bağlamda entropi-korelasyon matrisi Tablo 3.8'de gösterilmektedir.

Tablo 3. 8 Entropi- Korelasyon Matrisi

	1	2	3	4	5	6	7	8	9
1	1.00000	0.13110	0.03670	0.16343	-0.20953	0.02072	0.12739	0.02144	-0.20953
2	0.13110	1.00000	-0.14976	0.22976	-0.27336	-0.10754	0.09559	-0.09307	-0.27336
3	0.03670	-0.14976	1.00000	-0.20145	-0.24017	-0.01110	-0.02150	0.10495	-0.24017
4	0.16343	0.22976	-0.20145	1.00000	-0.29959	-0.19006	0.24392	-0.13526	-0.29959
5	-0.20953	-0.27336	-0.24017	-0.29959	1.00000	-0.35093	-0.07466	-0.15029	-0.05301
6	0.02072	-0.10754	-0.01110	-0.19006	-0.35093	1.00000	-0.08959	0.11875	-0.35093
7	0.12739	0.09559	-0.02150	0.24392	-0.07466	-0.08959	1.00000	0.03768	-0.07466

Tablo 3. 8 Entropi- Korelasyon Matrisi (devamı)

8	0.02144	-0.09307	0.10495	-0.13526	-0.15029	0.11875	0.03768	1.00000	-0.15029
9	-0.20953	-0.27336	-0.24017	-0.29959	-0.05301	-0.35093	-0.07466	-0.15029	1.00000

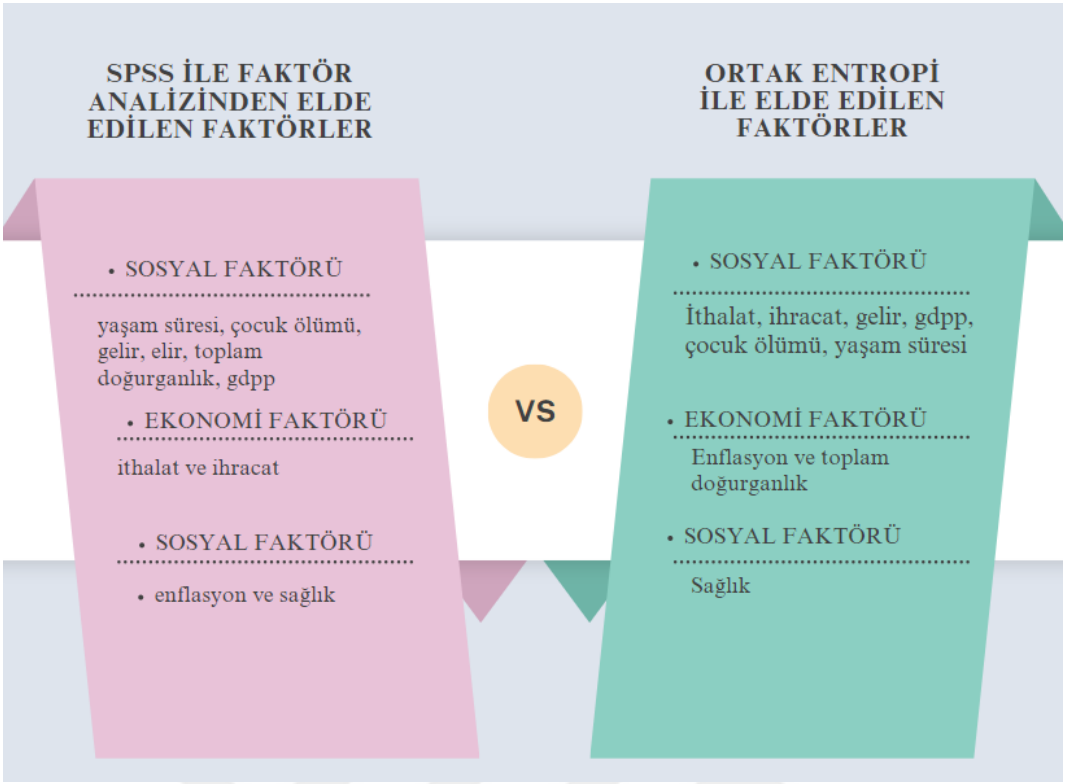
Entropi-korelasyon matrisinin elde edilmesinin ardından ise 9 değişkenden kaç tane faktöre indirgeneceği bulunacaktır. Faktör analizinde de olduğu gibi buradaki işlemlerde de Varimax dönüşümü kullanılmıştır. Mevcut 9 adet değişkenin 3 faktör ile temsil edilebileceğine karar verilir. Bu 3 faktör Tablo 3.9'da gösterilmiştir. Bu faktörlerin isimlendirilmesi ise değişkenlere bağlı olarak yapılmıştır.

Tablo 3. 9 Ortak entropi ile elde edilen faktörler

	Sosyal	**Ekonomi**	**Sağlık**
ithalat	0.765	-0.494	0.070
ihracat	0.613	-0.287	-0.018
gelir	-0.603	-0.255	-0.005
gdpp	-0.603	-0.255	-0.005
cocuk_olumu	0.539	0.005	0.072
yasam_suresi	0.340	-0.270	0.115
enflasyon	0.332	0.767	-0.566
toplam_dogurganlik	0.034	0.443	-0.018
saglik	0.090	0.693	0.720

Faktör 1'i temsilen toplum faktörü olarak isimlendirilmiştir. Sosyal faktöründeki değişkenler ise sırasıyla ithalat, ihracat, gelir, ithalat, gdpp (kişi başına düşen gayri safi yurtiçi hasılat), çocuk ölümü ve beklenen yaşam süresi şeklindedir. Faktör 2'yi temsilen ekonomi faktörü ele alınmıştır. Bahsi geçen faktördeki değişkenler ise enflasyon ve toplam doğurganlık değişkenleridir. Son olarak faktör 3'ü temsilen ise sağlık faktöründe sadece sağlık değişkeni bulunmaktadır.

Sonuç olarak yapılan bu iki uygulama sonrasında elde edilmiş olan faktörlerin karşılaştırılması Şekil 3.21'de gösterilmiştir.



Şekil 3.21 Faktör Analizi ve Ortak Entropi yöntemlerinden elde edilen faktörlerin karşılaştırılması

4
SONUÇ

4.1 Sonuç ve Öneriler

Çalışma kapsamında öncelikli hedef, istatistiksel anlamda entropi bazlı yaklaşımların da uygulamalarda bir alternatif olarak kullanılabileceğinin gösterilmesi hedeflenmiştir. Bu bağlamda ilk olarak entropi kavramı hakkında bilgiler verilmiş ve ortak entropi konusu ele alınmıştır. Bu bağlamda değişkenler arası ilişki entropi ölçüsü ile incelenmiştir ve bu yöntem ile birleşik entropi matrisi kullanılarak boyut indirgemenin sağlanabileceği incelenmiştir. Bu bilgilere ek olarak bibliyometik analiz üzerinde durulmuş ve bu amaç doğrultusunda yapılmış olan çalışmalara ait yazarlar, anahtar kelimeler ve literatür bilgileri kazanılmıştır.

Uygulama bölümünde ülkeler bazında genel gelişimini belirleyen sosyo-ekonomik ve sağlık faktörtörlerini kullanarak katagorize edilmiş değişkenlerin bulunduğu veri seti ile farklı analiz yöntemleri kullanılarak karşılaştırma yapılmıştır. İlk olarak veri seti faktör analizinin sonuçlarıyla entropi tabanlı yaklaşımın sonuçlarının karşılaştırması amacıyla faktör analizinin sonuçları elde edilmiştir. Ardında ise değişkenlere ait entropi ölçüleri ile elde edilen birleşik entropi matrisi yardımıyla boyut indirgeme sağlanıp sağlanmadığı teyit edilmiştir. Bu bilgiler doğrultusunda elde edilen bilgiler ile faktör analizinden elde edilen bilgilerin karşılaştırılması yapılmıştır.

Kullanılan veri setiyle entropi yöntemi sayesinde elde edilen birleşik entropi matrisi ile faktör indirgeme yöntemi gerçekleşitirilmesi hedeflenmiştir. Bu sayede birleşik entropi yöntemiyle faktör indirgemenin sağlanıp sağlanmadığı araştırılmıştır.

Sonuç olarak bu çalışma, entropinin teorik bir temelini ve varyans-kovaryans ölçümlerini kullanmanın yararlı olduğu alanlara ek olarak katkıda bulunmuştur. Bu çalışmanın, entropinin önemini ve entropiye dayalı araştırma yöntemlerini vurguladığı düşünüldüğünde, istatistik alanında entropi konusuna daha fazla önem verilmesi ve bu alanda daha fazla çalışma yapılması önerilmektedir. Bununla birlikte, çalışmanın ortaya çıkan zorluklar, akademik alanda yeni çalışmaların gerekli olduğunu göstermektedir. Çözülmesi gereken bazı sorunlara rağmen, alternatif olarak bilinen yöntemlerin yerine önerilmektedir.

KAYNAKÇA

[1] N. Ömürbek, G. Kılınç, and M. Karaatlı, "Entropi temelli Lanchester Savaş Modeli ile bir futbol maçının analizi," *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 14, no. 3, pp. 800–821, Jul. 2021, doi: 10.25287/ohuiibf.776905.

[2] Y. B. Altaylıgil, "Entropi Ölçüsü ve bazı Ekonometri Uygulamaları," *acikbilim*, 2008.

[3] O. Z. Çetinkaya, "Belirsizliğin Ölçülmesi ve Entropi,", İstanbul Üniversitesi İktisat Fakültesi Mecmuası, 44(1-4), 1986.

[4] C. Taslaman, "Din felsefesi açısından entropi yasası., (30), 3-25.," *Marmara Üniversitesi İlahiyat Fakültesi Dergisi*, no. 30, pp. 3–25, 2006.

[5] Ş. A. Baray, "Entropi ve Karar Verme," *Yönetim Dergisi: İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadı Enstitüsü*, vol. 14, no. 44, pp. 7–21, 2003.

[6] S. Kullback, *Information Theory and Statistics*. 1960, Courier Corporation.

[7] Y. Polat, "Faktör Analizi Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi ve Hayvancılık Denemesine Uygulanışı," 2012.

[8] K. Özdamar, "Paket Programlar ile İstatistiksel Veri Analizi," *Ankara: Nisan Kitapevi, (Cilt 1)*, pp. 27–36, 2013.

[9] Ş. Büyüköztürk, " Faktör analizi: Temel kavramlar ve Ölçek Geliştirmede Kullanımı," *Kuram ve uygulamada eğitim yönetimi*, vol. 32, no. 32, pp. 470–483, 2002.

[10] H. Tatlıdil, "Uygulamalı Çok Değişkenli İstatistiksel Analiz.," *Engin Yayınları*, 1992.

[11] M. İnan, "Sosyo-Ekonomik Gelişmişlik Siralamasi Açisindan Ordu İli (Ülke Geneli ve Karadeniz Bölgesi Bir Karşılaştırma)," vol. 6, no. 4, pp. 76–103, 2020.

[12] C. E. Shannon, "A Mathematical Theory Of Communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[13] M. A. Koşan, A. Coşkun, and H. Karacan, " Yapay Zeka Yöntemlerinde Entropi," *Journal of Information Systems and Management Research*, vol. 1, no. 1, pp. 15–22, 2019.

[14] A. B. Cambel, " Applied chaos theory: A Paradigm For Complexity. ," *Elsevier*, 1993.

[15] Esra Öncül, "Çok Değişkenli Yöntemlerde Entropi Kullanımı: Mutluluk İndeksi Üzerine Bir Uygulama," Yıldız Teknik Üniversitesi, 2023.

[16] K. S. Wairegi, "Shannon Mathematical Theory Of Communication".

[17] İ. Değirmenci, "Entropi Ölçümleri ve Maksimum Entropi İlkesi (Doctoral dissertation, Yüksek Lisans Tezi, Hacettepe Üniversitesi).," Yüksek Lisans Tezi, Hacettepe Üniversitesi, 2011.

[18] O. Bahadır and H. Türkmençalıkoğlu, "Bilgi Kuramında Shannon Entropisi ve Uygulamaları," *Avrupa Bilim ve Teknoloji Dergisi*, no. 32, pp. 497–497, 2021.

[19] T. M. Cover, "Elements Of İnformation Theory," *John Wiley & Sons*, 1999.

[20] S. Çağdaş and B. Karaçalı, "Ortak Bilgi Miktarının Modelden-Bağımsız ve Hızlı Hesaplanması İçin Yeni Yöntemler," *In 26th Signal Processing and Communications Applications Conference, SIU 2018. Institute of Electrical and Electronics Engineers Inc.*., 2018.

[21] T. M. Cover and J. A. Thomas, "Entropy, Relative Entropy and Mutual Information," *Elements of information theory*, vol. 2, no. 1, pp. 12–13, 1991.

[22] E. Çöp, "Faktör analizi ve kümeleme analizinin karşılaştırılması: Teori ve uygulama," İnönü Üniversitesi Sosyal Bilimler Enstitüsü, Master's thesis, 2022.

[23] B. Ergül, "Sağlam faktör analizi ve bir uygulama," Doktora tezi, Eskişehir Osmangazi Üniversitesi, 2023.

[24] F. İlhan, "Faktör Analizi ve Tarımsal Araştırmalarda Elde Edilen Verilere Uygulanması Üzerine Bir Çalışma," Master's thesis, Selçuk Üniversitesi, 2007.

[25] R. A. Johnson and D. W. Wichern, “Applied multivariate statistical analysis,” 2002.

[26] S. Patır, “Faktör Analizi İle Öğretim Üyesi Değerleme Çalişmasi,” *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 23, no. 4, pp. 69–86, 2009.

[27] E. Tavşancıl, “Tutumların Ölçülmesi Ve SPSS İle Veri Analizi,” *Nobel Yayınları, Ankara*, 2010.

[28] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, and R. Tatham, “Multivariate Data Analysis,” *Uppersaddle River*, 2006.

[29] M. Turanlı, D. Cengiz, and Ö. Bozkır, “Faktör Analizi İle Üniversiteye Giriş Sinavlarindaki Başari Durumuna Göre İllerin Sıralanmasi,” *Istanbul University Econometrics and Statistics e-Journal*, no. 17, pp. 45–68, 2012.

[30] Ş. Kalaycı, “SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri,” *Asil Yayın Dağıtım, Ankara, Turkey*, vol. 5, no. 359, 2010.

[31] N. B. Çilingir, “Faktör Analizi: Teori ve Uygulama ,” Master’s thesis, Sosyal Bilimler Enstitüsü, 2019.

[32] M. S. Lewis-Beck, “Factor Analysis And Related Techniques,” *Sage Publications, Singapore*, 1994.

[33] S. Süzülmüş, “Faktör analizi modellerinin belirlenebilirliği ve genelleştirilmiş inverslerin kullanımı,” Yüksek Lisans Tezi, Çukurova Üniversitesi, 2005.

[34] J. F. Hair, R. E. Anderson, R. L. Tatham, and W. C. Black, “Multivariate Data Analysis With Readings,” *New York, NY: Pearson Prentice Hall.*, vol. 6, 2006.

[35] A. S. Albayrak, “Türkiye’de İllerin Sosyoekonomik Gelişmişlik Düzeylerinin Çok Değişkenli İstatistik Yöntemlerle İncelenmesi,” *Uluslararası Yönetim İktisat ve İşletme Dergisi*, vol. 1, no. 1, pp. 153–177, 2005.

[36] Y. Toktay, “Çok Değişkenli İstatistik Analiz Yöntemler: Faktör Analizi ve Diskriminant Analizinin Iğdır Üniversitesi Öğrencileri Üzerine Uygulaması,” Master’s thesis, Fen Bilimleri Enstitüsü, 2017.

[37] A. L. Comrey, “Common Methodological Problems In Factor Analytic Studies,” *J Consult Clin Psychol*, vol. 46, no. 4, p. 648, 1978.

[38] E. Özgür, “Çok Değişkenli İstatistiksel Analiz Yöntemleri Ve Bir Uygulama," Doktora Tezi, , 2003,” Gazi Üniversitesi, Ankara, 2003.

[39] Ş. Büyüköztürk, “Faktör Analizi: Temel Kavramlar ve Ölçek Geliştirmede Kullanımı,” *Kuram ve uygulamada eğitim yönetimi*, vol. 32, no. 32, pp. 470–483, 2002.

[40] Y. Wei, “Variance, Entropy, and Uncertainty Measure ,” *Dept. of Stastistics, People’s Univ. of China*, 1987.

[41] I. Virtanen and J. Astola, “On entropy-based dependence measures for two and three dimensional categorical variable distributions,” *In Proceedings of The First World Congress of Bernoulli Society for Mathematical Statistics and Probability*, 1986.

[42] Ü. Erbaş, “Entropi İlkelerinin Boyut İndirgeme Uygulamaları,” Doctoral dissertation, Marmara Üniversitesi, 2010.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. N. ORUÇ and D. YILDIZ, "İstatistiksel Analizde Ortak Entropi Uygulaması," presented at the 7th International Researchers, Statisticians, and Young Statisticians Congress , 2023.