



MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ

Fen Edebiyat Fakültesi

İstatistik Bölümü

Çok Değişkenli İstatistiksel Yöntemler

Dönem Sonu Projesi

Adı Soyadı: Ahmet Can GÜNAY

Öğrenci Numarası: 20242107004

Table of Contents

1. Veri Seti Özeti	4
1.1 Veri Setindeki Nicel Değişkenler	4
1.2 Veri Setindeki Kategorik Değişkenler	4
2. Korelasyon Matrisi ve Değişken Seçimi.....	6
3. Keşifçi Veri Analizi (EDA).....	8
3.1 Kutu (Boxplot) Grafikleri.....	9
3.2 Genel Özet İstatistikleri (Data Summary)	10
3.3 Çarpıklık ve Basıklık Katsayıları Analizi (Skewness & Kurtosis Analysis)	11
4. Normallik Analizi	15
4.1 Q-Q Plot Analizi.....	15
4.2 Normallik Testleri (Shapiro-Wilk ve Kolmogorov-Smirnov Testleri).....	17
4.3 Normallik Dönüşümleri	19
5. MANOVA Analizi	22
5.1 Tek Yönlü MANOVA Analizi.....	22
5.2 Çift Yönlü MANOVA Analizi	32
6. Temel Bileşenler ve Faktör Analizi	43
7. Diskriminant Analizi	57
8. Lojistik Regresyon Analizi	71
9. Kümeleme Analizi.....	78

GİRİŞ

Bu çalışma, sađlık ve fitness verilerinin analizi yoluyla kiřiselleřtirilmiř sađlık önerileri geliřtirmeyi amaçlayan bir arařtırma projesidir. Günümüzde giyilebilir teknolojiler ve akıllı cihazlar aracılıđıyla toplanan sađlık verilerinin analizi, bireylerin sađlık durumlarını iyileřtirmede ve yařam kalitelerini artırmada önemli bir rol oynamaktadır.

Bu projede, bir yıllık süre boyunca 3.000 katılımcıdan bilgi içeren kapsamlı bir sađlık ve zindelik izleme veri seti olan FitnessApp¹ veri setini analiz eder. Bu veri seti 2023 yılında İstanbul Kültür Üniversitesi Elektrik Elektronik Mühendisliđi bitirme projesi kapsamında mobil uygulama ve akıllı saat geliřtirme çalışmalarının ardından bu cihazlar üzerinden verilerin toplanması ile oluşturulmuřtur. Veri seti, 18-65 yař aralıđındaki çeřitli bir popölasyonda fiziksel aktivite, sađlık ölçümleri ve yařam tarzı faktörleri arasındaki iliřkiye dair deđerli içgörüler sađlar.

Arařtırmanın konusu, FitApp veri setindeki katılımcıların fitness seviyesi kategorilerine (Low, Medium, High) göre; yař, cinsiyet ve sađlık durumu gibi demografik özellikleri ile günlük adım sayısı, uyku süresi, stres seviyesi ve hidrasyon seviyesi gibi yařam tarzı metriklerinin deđiřip deđiřmediđidir. Bu özelliklerin ve metriklerin farklı fitness seviyelerindeki deđiřimini üç farklı program kullanarak (SAS, R ve SPSS) en uygun istatistiksel analizlerle incelemek ve sonuçları karřılařtırmalı olarak ortaya koymaktır.

¹ <https://github.com/ahmetcangunay/fitness-app-data>

1. Veri Seti Özeti

Veri seti 3000 katılımcıdan oluşan ve bir yıllık süre boyunca toplanan sağlık ve fitness verilerini içermektedir. Veri seti toplamda 21 değişkenden oluşmaktadır. Analizde daha anlamlı sonuçlar elde edilebilmesi adına `fitness_level` değişkeni üç kategoriye (Low, Medium, High) ayrılmıştır ve `fitness_level_category` olarak yeniden düzenlenmiştir.

1.1 Veri Setindeki Nicel Değişkenler

Veri setindeki nicel değişkenler ve açıklamaları aşağıda verilmiştir:

1. **`fitness_level`:** Katılımcıların genel fitness seviyesi puanı (bağımlı değişken)
2. **`age`:** Katılımcıların yaşı (18-65 yaş aralığı)
3. **`daily_steps`:** Günlük atılan adım sayısı
4. **`calories_burned`:** Günlük yakılan kalori miktarı
5. **`duration_minutes`:** Aktivite süresi (dakika)
6. **`avg_heart_rate`:** Aktivite sırasındaki ortalama kalp atış hızı
7. **`resting_heart_rate`:** Dinlenme halindeki kalp atış hızı
8. **`blood_pressure_systolic`:** Sistolik kan basıncı değeri
9. **`blood_pressure_diastolic`:** Diastolik kan basıncı değeri
10. **`hours_sleep`:** Günlük uyku süresi (saat)
11. **`stress_level`:** Günlük stres seviyesi (1-10 arası)
12. **`hydration_level`:** Günlük su tüketimi (litre)
13. **`weight_kg`:** Vücut ağırlığı (kg)
14. **`height_cm`:** Boy uzunluğu (cm)
15. **`bmi`:** Vücut kitle indeksi (Vücut ağırlığı ve boy uzunluğu üzerinden hesaplanmıştır.)

1.2 Veri Setindeki Kategorik Değişkenler

Veri setindeki kategorik değişkenler ve açıklamaları aşağıda verilmiştir:

1. **`fitness_level_category`:** Fitness seviyesi kategorisi ('Low', 'Medium', 'High') (Kategorik hale getirilmiş bağımlı değişken)
2. **`gender`:** Cinsiyet ('M', 'F', 'Other')
3. **`activity_type`:** Aktivite türü ('Tennis', 'Cycling', 'Yoga', 'Weight Training', 'Dancing', 'Running', 'HIIT', 'Walking', 'Basketball', 'Swimming')

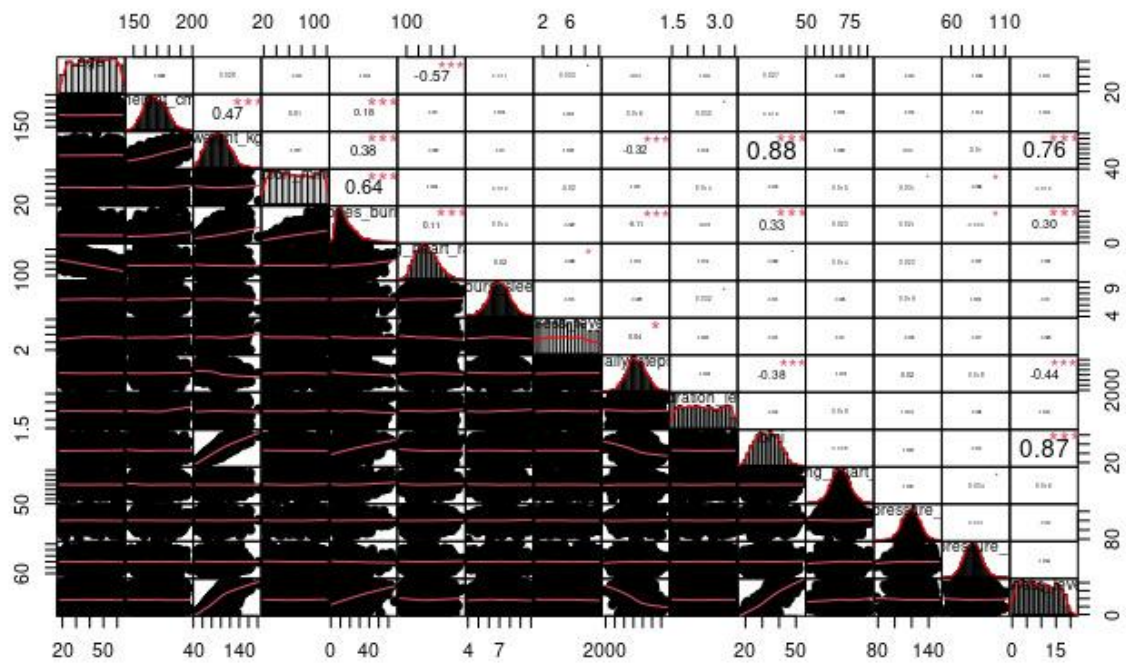
4. **intensity:** Egzersiz yoğunluđu ('Low', 'Medium', 'High')
5. **smoking_status:** Sigara kullanım durumu ('Never', 'Former', 'Current')
6. **health_condition:** Sađlık durumu ('Healthy', 'Diabetes', 'Hypertension', 'Asthma')

Veri setindeki toplam 3000 gözlem ve 21 deđişken üzerinde gerçekleştirilen keşifsel veri analizi kapsamında, deđişkenler arası ilişkileri incelemek ve optimal deđişken seçimini gerçekleştirmek amacıyla korelasyon matrisi analizi uygulanmıştır. Bu ön eleme aşaması, deđişkenler arasındaki çoklu doğrusal bağlantı problemini önlemek ve analiz verimliliđini artırmak için gerçekleştirilmiştir.

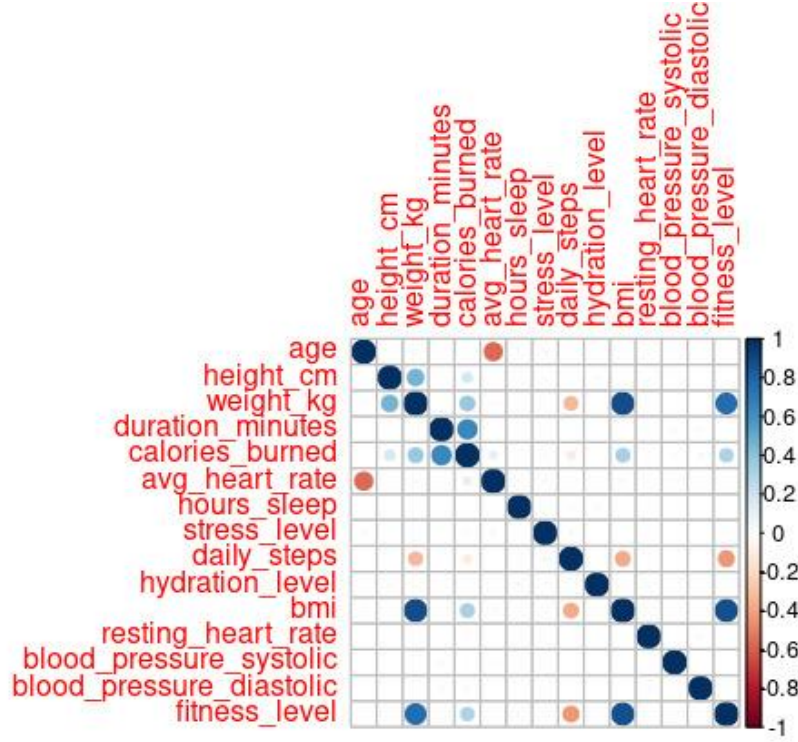
2. Korelasyon Matrisi ve Değişken Seçimi

Bu bölümde, veri setindeki değişkenler arasındaki ilişkiler incelenmiş ve çoklu bağlantı problemlerini azaltmak amacıyla gereksiz değişkenlerin çıkarılması hedeflenmiştir.

Veri setindeki nicel değişkenler arasında bir korelasyon matrisi oluşturulmuş, değişkenler arasındaki ilişkiler ölçülmüştür. Korelasyon katsayılarının incelenmesiyle, $|korelasyon\ katsayısı| > 0.7$ olan değişkenler belirlenmiştir.



Şekil 1 Korelasyon Matrisi



Şekil 2 Korelasyon Matrisi 2

Şekil 1 ve Şekil 2'deki korelasyon matrisleri incelendiğinde fitness_level ve BMI arasında 0.87'lik pozitif güçlü bir bağlantı olduğu gözlemlenmiştir. Aynı şekilde fitness_level ve weight_kg arasında da 0.76'lık pozitif bir ilişki olduğu gözlemlenmektedir. BMI (vücut kitle endeksi) boy ve kilo üzerinden hesaplandığından boy ve kilo değişkenleri veri setinden çıkartılmıştır. Fitness_level ve BMI arasındaki durum, BMI değişkeninin fiziksel aktivitelerde oldukça önemli olduğu bilindiğinden gözardı edilip analize devam edilmiştir.

3. Keşifçi Veri Analizi (EDA)

Gerçekleştirilen değişken seçimi ve veri ön işleme aşamasının ardından, başlangıçta 21 değişken içeren veri setindeki özellik sayısı 19'a indirgenmiştir. Veri setinin gözlem sayısı ise 3000 olarak korunmuştur, böylece analizler için yeterli örneklem büyüklüğü sağlanmıştır.

```
> str(fitness_app_data)
tibble [3,000 × 19] (S3: tbl_df/tbl/data.frame)
 $ age          : num [1:3000] 56 40 23 37 50 61 22 54 20 48 ...
 $ gender       : chr [1:3000] "F" "M" "F" "F" ...
 $ activity_type : chr [1:3000] "Tennis" "Cycling" "Yoga" "Yoga" ...
 $ duration_minutes : num [1:3000] 31 91 90 101 71 98 36 20 32 109 ...
 $ intensity     : chr [1:3000] "Low" "Low" "Medium" "Medium" ...
 $ calories_burned : num [1:3000] 5.2 27.7 10.5 9.4 22.3 12.7 6.3 7 10.6 26.5 ...
 $ avg_heart_rate : num [1:3000] 104 119 157 143 123 108 157 131 133 108 ...
 $ hours_sleep   : num [1:3000] 6.9 7.5 4.4 6.4 6.5 6.9 7.9 8.2 6.4 9.1 ...
 $ stress_level  : num [1:3000] 9 5 3 8 4 9 7 4 6 2 ...
 $ daily_steps   : num [1:3000] 10977 5494 6696 10457 7743 ...
 $ hydration_level : num [1:3000] 1.8 2.5 2.4 2.2 1.7 1.8 2.7 3.3 2.6 2.2 ...
 $ bmi          : num [1:3000] 30.5 41 46.5 33.5 37.5 ...
 $ resting_heart_rate : num [1:3000] 69.5 82.1 68 60.7 70.2 68.7 74.6 72.2 65.5 66.9 ...
 $ blood_pressure_systolic : num [1:3000] 111 129 118 130 139 ...
 $ blood_pressure_diastolic : num [1:3000] 72.9 78.6 69 69.3 80.7 78 89.3 89.2 94.7 69 ...
 $ health_condition : chr [1:3000] "Healthy" "Healthy" "Healthy" "Healthy" ...
 $ smoking_status  : chr [1:3000] "Never" "Never" "Never" "Never" ...
 $ fitness_level   : num [1:3000] 9.42 17.74 17.31 8.08 8.04 ...
 $ fitness_level_category : chr [1:3000] "Medium" "High" "High" "Medium" ...
```

Şekil 3 Veri Özeti

Veri setindeki değişkenlerin veri türleri, her bir değişkene ait gözlem sayıları ve örnek veri içerikleri Şekil 3'te detaylı olarak gösterilmektedir. Bu görsel çıktı, verilerin yapısal özelliklerini ve dağılımını incelemek için gerçekleştirilen keşifsel veri analizi aşamasının bir parçası olarak elde edilmiştir.

```
> colSums(is.na(fitness_app_data))
      age      gender      activity_type      duration_minutes      intensity      calories_burned      avg_heart_rate
      0           0           0           0           0           0           0
  hours_sleep      stress_level      daily_steps      hydration_level      bmi      resting_heart_rate      blood_pressure_systolic
      0           0           0           0           0           0           0
blood_pressure_diastolic      health_condition      smoking_status      fitness_level      fitness_level_category
      0           0           0           0           0
```

Şekil 4 Eksik Veri Analizi

Veri setindeki eksik değerlerin incelenmesi için gerçekleştirilen analiz sonucunda (Şekil 4), veri setinde herhangi bir eksik gözlem tespit edilmemiştir. Bu nedenle, veri ön işleme aşamasında eksik veri tamamlama yöntemlerinin uygulanmasına gerek duyulmamıştır.

```
> table(fitness_app_data$activity_type)
Basketball      Cycling      Dancing      HIIT      Running      Swimming      Tennis      Walking      Weight Training      Yoga
      299           288           311           292           296           307           298           283           324           302
```

Şekil 5 Aktivite Tipi Dağılımı


```
> table(fitness_app_data$intensity)

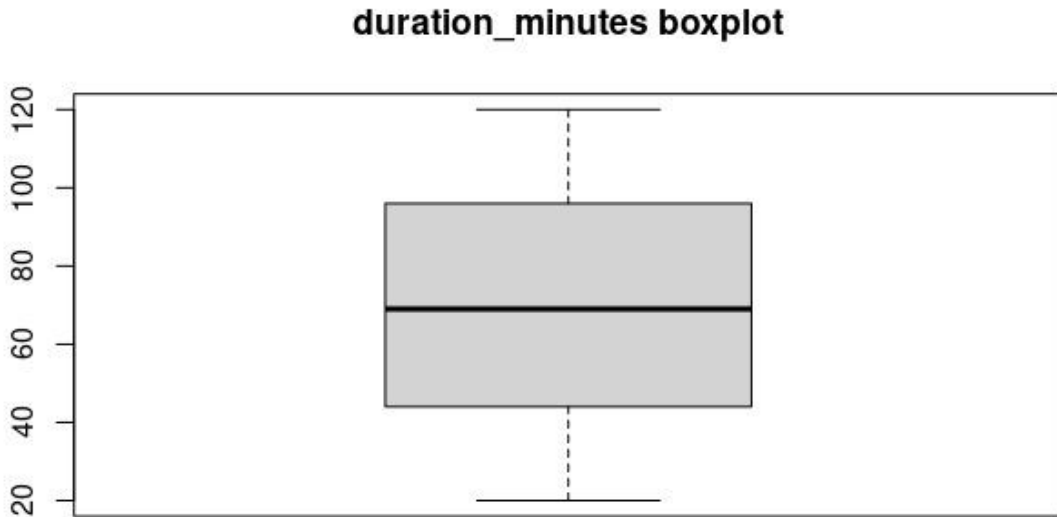
High   Low Medium 
  545   914  1541
```

Şekil 6 Aktivite Yoğunluğu Dağılımı

Şekil 5'te gösterilen aktivite türlerinin frekans dağılımları incelendiğinde, her bir aktivite kategorisinin örneklem içerisinde yaklaşık olarak eşit temsil edildiği ve istatistiksel analizler için dengeli bir dağılım sergilediği gözlemlenmiştir. Şekil 6'da sunulan aktivite yoğunluğu dağılımı analiz edildiğinde, düşük (low) ve orta (medium) yoğunluktaki aktivitelerin yüksek (high) yoğunluktaki aktivitelere kıyasla daha fazla frekansa sahip olduğu ve bu durumun örneklemde asimetrik bir dağılım oluşturacağı tespit edilmiştir.

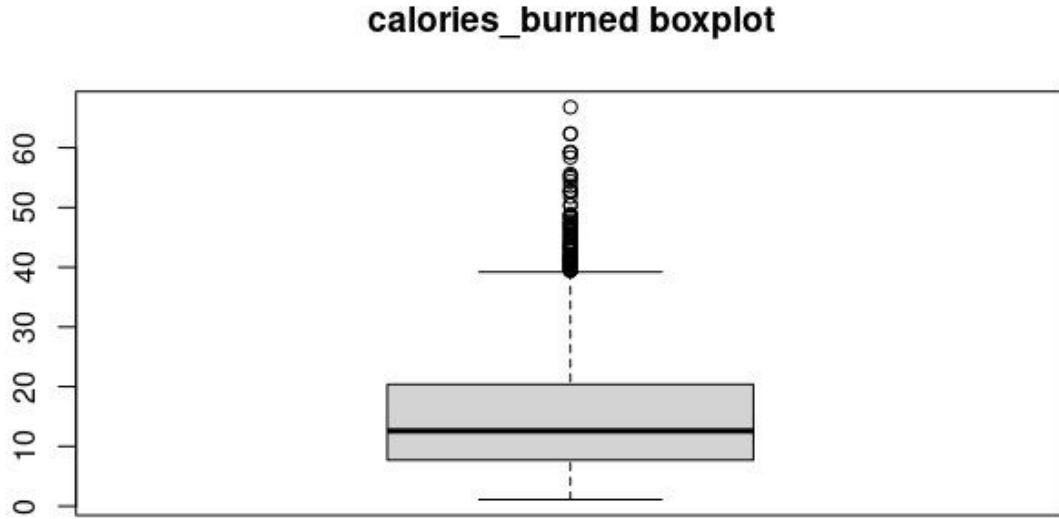
3.1 Kutu (Boxplot) Grafikleri

Veri setindeki potansiyel uç değerlerin tespit edilmesi amacıyla seçili değişkenler için kutu grafikleri oluşturulmuştur. Bu görsel analiz yöntemi, veri dağılımındaki aykırı gözlemlerin belirlenmesine olanak sağlamıştır.



Şekil 7 Aktivite Süresi Kutu Grafiği

Şekil 7’de sunulan aktivite süresine ait kutu grafiği incelendiğinde, verilerin dağılımının normal bir dağılım sergilediği gözlemlenmektedir.



Şekil 8 Yakılan Kalori Kutu Grafiği

Şekil 8’de sunulan yakılan kaloriye ait kutu grafiği incelendiğinde, uç değerlerin belirli bir aralıkta yoğunlaştığı gözlemlenmektedir. Bu durum, veri setinde olağan dışı değerlerin bulunduğunu ve bu değerlerin belirli bir bölgeye odaklandığını göstermektedir. Uç değerlerin analizi, verilerin geneli üzerindeki etkisini değerlendirmek ve potansiyel sebeplerini belirlemek için daha detaylı bir inceleme gerektirebilir.

3.2 Genel Özet İstatistikleri (Data Summary)

Şekil 9’da yer alan fitness uygulaması verilerinin özet istatistiklerine göre, katılımcıların yaşları 18 ile 64 arasında değişmekte olup, yaş ortalaması 41.64 olarak hesaplanmıştır. Veride cinsiyet ve aktivite türü kategorik değişkenler olarak tanımlanmıştır. Aktivite süresi 20 ile 120 dakika arasında değişirken, medyan değeri 69 dakika olarak gözlenmiştir. Buna paralel olarak, yakılan kalori miktarı minimum 1.10 ile maksimum 66.80 arasında değişmekte, medyan değeri ise 12.60’tır.

Katılımcıların fiziksel sağlık göstergelerine bakıldığında, BMI (Vücut Kitle İndeksi) değerlerinin 16.84 ile 54.29 arasında değiştiği, medyan değer ise 32.92 olduğu görülmektedir. Dinlenme kalp atış hızı ortalama 70 iken, sistolik kan basıncı 78 ile 152.7 arasında, diyastolik kan basıncı ise 53.7 ile 112.1 arasında değişmektedir. Günlük yaşam tarzı verileri incelendiğinde, katılımcıların günlük adım sayısının 1664 ile 15227 arasında değiştiği ve medyan değerinin 8564 olduğu dikkat çekmektedir. Uyku süresi ise 4 ile 10 saat arasında değişmekte olup, ortalama 7 saat olarak belirlenmiştir. Ayrıca, günlük hidrasyon seviyeleri 1.5 ile 3.5 litre arasında değişmekte ve medyan değer 2.5 litredir.

Son olarak, fitness seviyesi 0.04 ile 21.64 arasında değişmekte olup, medyan değer 9.905'tir. Bu seviye, fitness seviyesine göre sınıflandırılmış kategorik bir değişkenle de desteklenmiştir. Genel olarak, bu özet istatistikler, katılımcıların fiziksel aktivite ve sağlık durumları hakkında genel bir çerçeve sunmakta ve verinin daha ileri analizler için uygun olduğunu göstermektedir.

```
> summary(fitness_app_data)
```

age	gender	activity_type	duration_minutes	intensity	calories_burned	avg_heart_rate	hours_sleep
Min. :18.00	Length:3000	Length:3000	Min. : 20.00	Length:3000	Min. : 1.10	Min. : 89.0	Min. : 4.000
1st Qu.:30.00	Class :character	Class :character	1st Qu.: 44.00	Class :character	1st Qu.: 7.70	1st Qu.:118.0	1st Qu.: 6.400
Median :42.00	Mode :character	Mode :character	Median : 69.00	Mode :character	Median :12.60	Median :130.0	Median : 7.000
Mean :41.64			Mean : 69.87		Mean :15.05	Mean :131.3	Mean : 7.015
3rd Qu.:53.00			3rd Qu.: 96.00		3rd Qu.:20.40	3rd Qu.:144.0	3rd Qu.: 7.700
Max. :64.00			Max. :120.00		Max. :66.80	Max. :194.0	Max. :10.000
stress_level	daily_steps	hydration_level	bmi	resting_heart_rate	blood_pressure_systolic	blood_pressure_diastolic	health_condition
Min. : 1.000	Min. : 1664	Min. :1.500	Min. :16.84	Min. :51.10	Min. : 78.0	Min. : 53.70	Length:3000
1st Qu.: 3.000	1st Qu.: 7279	1st Qu.:2.000	1st Qu.:27.59	1st Qu.:66.50	1st Qu.:113.1	1st Qu.: 74.60	Class :character
Median : 5.000	Median : 8564	Median :2.500	Median :32.92	Median :70.00	Median :120.2	Median : 80.10	Mode :character
Mean : 5.284	Mean : 8670	Mean :2.491	Mean :33.01	Mean :70.01	Mean :120.0	Mean : 80.19	
3rd Qu.: 8.000	3rd Qu.:10039	3rd Qu.:3.000	3rd Qu.:38.40	3rd Qu.:73.50	3rd Qu.:127.0	3rd Qu.: 85.70	
Max. :10.000	Max. :15227	Max. :3.500	Max. :54.29	Max. :87.10	Max. :152.7	Max. :112.10	
smoking_status	fitness_level	fitness_level_category					
Length:3000	Min. : 0.040	Length:3000					
Class :character	1st Qu.: 4.380	Class :character					
Mode :character	Median : 9.055	Mode :character					
	Mean : 9.267						
	3rd Qu.:14.020						
	Max. :21.640						

Şekil 9 Özet İstatistikler

3.3 Çarpıklık ve Basıklık Katsayıları Analizi (Skewness & Kurtosis Analysis)

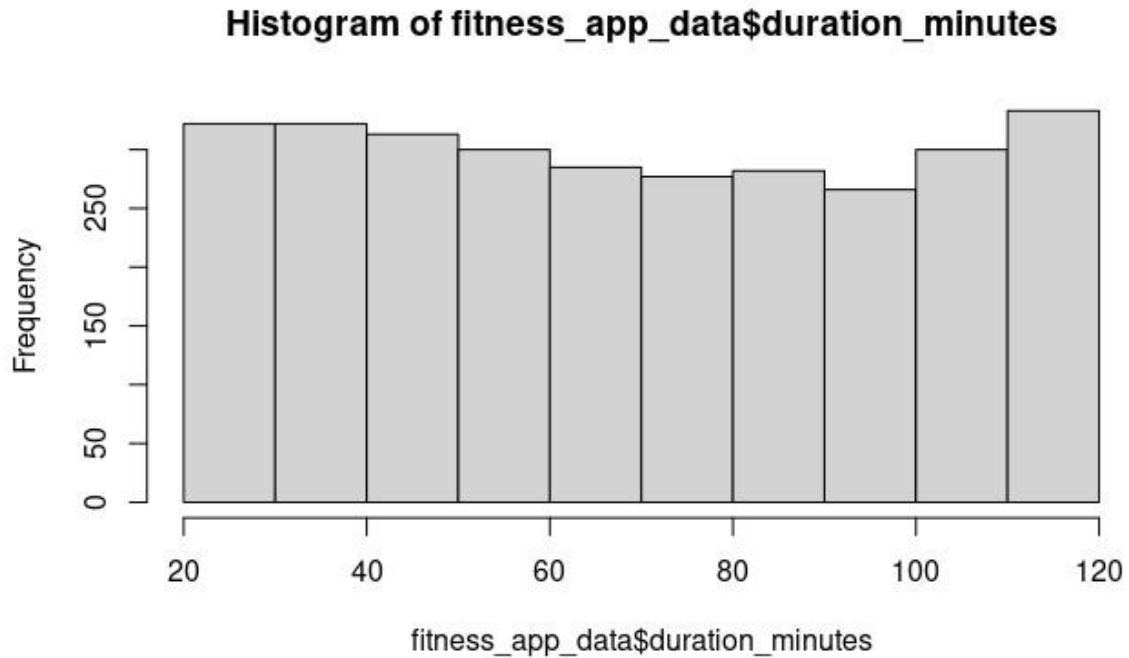
Bölüm 3.1'de incelenen değişkenler olan aktivite süresi ve yakılan kalori miktarı, çarpıklık ve basıklık katsayıları üzerinden değerlendirilecektir. Çarpıklık (skewness) katsayısı, dağılımın simetrik olup olmadığını belirlerken, basıklık (kurtosis) katsayısı ise dağılımın tepe noktasının keskinliği veya düzlüğü hakkında bilgi vermektedir.

```
> skewness(fitness_app_data$duration_minutes)
[1] 0.03857828
> kurtosis(fitness_app_data$duration_minutes)
[1] 1.760953
```

Şekil 10 Aktivite Süresinin Çarpıklık ve Basıklık Değerleri

Şekil 10'da elde edilen sonuçlara göre, aktivite süresi değişkeninin çarpıklık (skewness) katsayısı 0.0386 olarak hesaplanmıştır. Bu değer, dağılımın neredeyse simetrik olduğunu ve çarpıklığın ihmal edilebilir düzeyde olduğunu göstermektedir. Basıklık (kurtosis) katsayısı ise 1.761 olarak bulunmuştur, bu da dağılımın normal dağılıma kıyasla daha düz (platykurtik) bir yapıya sahip olduğunu ifade etmektedir.

Aktivite süresi değişkeni normal dağılıma oldukça yakın bir dağılım sergilemektedir. Ancak, basıklık katsayısının düşük olması, verinin uç değerler açısından daha az yoğun olduğunu göstermektedir. Bu durum, aktivite süresi verisinin analizinde uç değerlere karşı daha az duyarlı bir dağılıma işaret etmektedir.



Şekil 11 Aktivite Süresi Histogram Grafiği

Histogram grafiđi incelendiđinde (Şekil 11), aktivite süresi deđişkeninin dađılımını görselleştirmektedir. Grafik incelendiđinde, çubukların yüksekliđi genel olarak benzer olduđu için dađılımın simetriye yakın olduđu görölmektedir. Bu durum, Şekil 10'daki çarpıklık katsayısının (0.0386) düşük olmasıyla da uyumludur. Ancak, uç deđerlerin yoğun bir şekilde birikmediđi ve dađılımın oldukça eşit bir şekilde yayıldıđı gözlenmektedir, bu da bir önceki şekilde yer alan düşük basıklık katsayısına (1.761) işaret etmektedir.

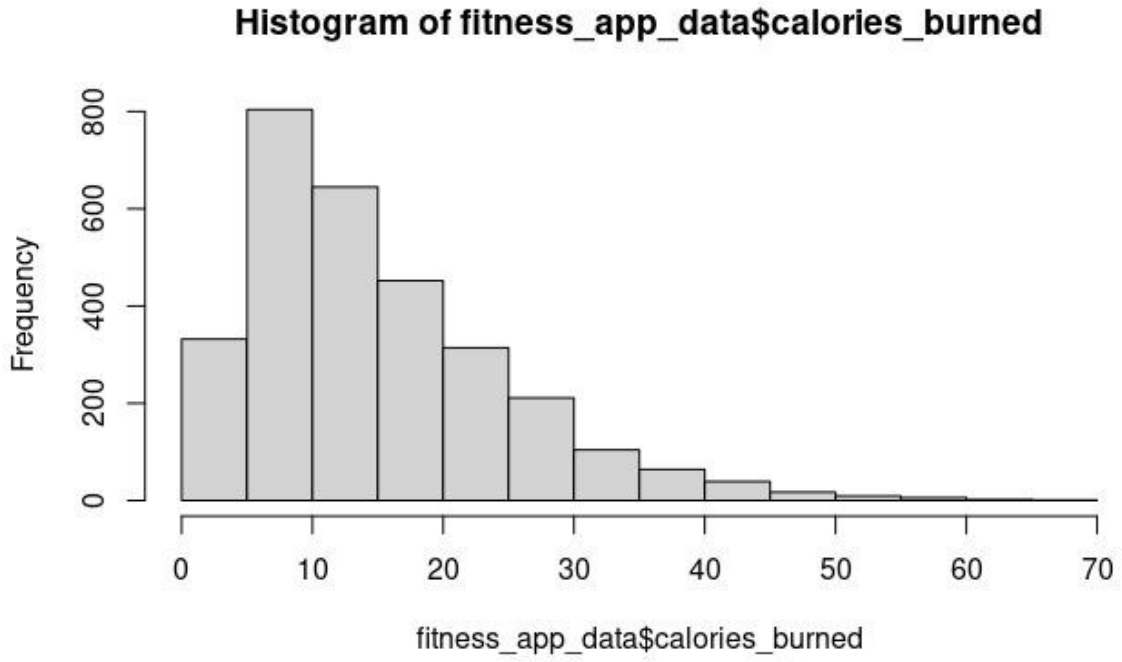
Tüm bu sonuçlar toplandıđında, histogram, aktivite süresi deđişkeninin yaklaşık olarak uniform (düzgün) bir dađılıma sahip olduđunu göstermekte ve çarpıklık ile basıklık analizinden elde edilen bulguları desteklemektedir.

```
> skewness(fitness_app_data$calories_burned)
[1] 1.270696
> kurtosis(fitness_app_data$calories_burned)
[1] 4.931033
```

Şekil 12 Yakılan Kalori Miktarının Çarpıklık ve Basıklık Deđerleri

Elde edilen bulgulara göre (Şekil 12), yakılan kalori miktarının çarpıklık (skewness) katsayısı 1.270696 olarak hesaplanmıştır. Bu deđer, dađılımın sađa çarpık olduđunu göstermektedir. Bu çarpıklık, veri noktalarının büyük bir kısmının düşük kalori aralıklarında yoğunlaştıđını, ancak daha az sayıda büyük kalori deđerlerinin de bulunduđunu ifade etmektedir.

Basıklık (kurtosis) katsayısı ise 4.931033 olarak bulunmuştur. Bu deđer, dađılımın normal dađılıma kıyasla daha sivri (leptokurtik) bir yapıya sahip olduđunu göstermektedir. Bu durum, veri kümesinin uç deđerler açısından daha yoğun bir yapı sergilediđine işaret etmektedir.



Şekil 13 Yakılan Kalori Miktarı Histogram Grafiği

Şekil 13'teki histogram grafiği incelendiğinde, yakılan kalori miktarı değişkeninin dağılımı görselleştirilmiştir. Grafikte, değerlerin düşük kalorilerde yoğunlaştığı ve sağ tarafa doğru azalan bir eğilim sergilediği görülmektedir. Bu gözlem, çarpıklık katsayısının sağa çarpık bir dağılıma işaret eden değeriyle uyumludur. Ayrıca, yüksek kurtosis değeri, dağılımda uç değerlerin bulunduğunu da desteklemektedir.

Sonuç olarak, histogram ve çarpıklık ile basıklık analizleri, yakılan kaloriler değişkeninin sağa çarpık ve uç değerler açısından yoğun bir dağılıma sahip olduğunu ortaya koymaktadır. Bu durum, analiz sırasında uç değerlerin dikkate alınması gerektiğini göstermektedir.

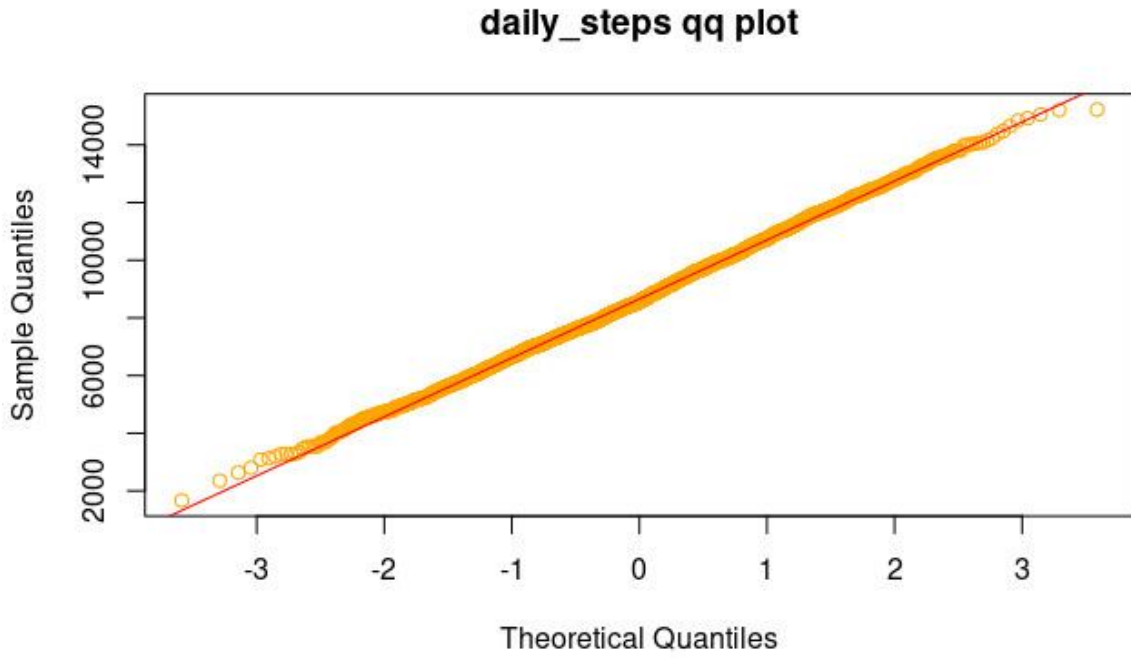
4. Normallik Analizi

Araştırmanın istatistiksel analiz sürecinin ilk aşaması olan normallik analizi, veri setindeki sürekli değişkenlerin dağılım karakteristiklerini incelemek ve uygun analiz yöntemlerinin seçimini belirlemek amacıyla gerçekleştirilmiştir. Bu bağlamda, katılımcıların günlük adım sayısı, uyku süresi, stres seviyesi ve hidrasyon seviyesi gibi metriklerinin normal dağılıma uygunluğu, görsel ve sayısal yöntemler kullanılarak değerlendirilmiştir. Normal dağılım varsayımının sağlanmadığı değişkenler için uygun dönüşüm teknikleri uygulanmış ve bu dönüşümlerin etkinliği, Q-Q plot, Shapiro-Wilk testi ve Kolmogorov-Smirnov testi gibi çeşitli normallik testleri aracılığıyla analiz edilmiştir.

4.1 Q-Q Plot Analizi

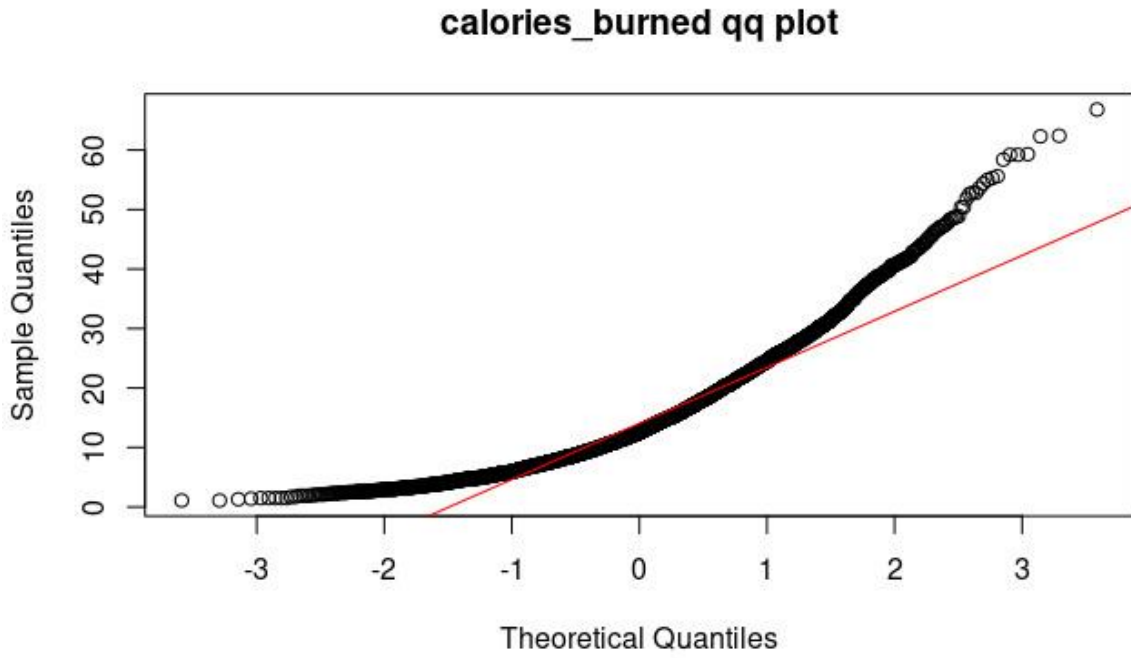
Bu çalışmada, FitnessApp veri setindeki değişkenlerin teorik normal dağılımla uyumunu değerlendirmek ve analiz öncesinde dağılım özelliklerini görselleştirmek için Q-Q plot kullanılmıştır. Fitness seviyesi, yaş, uyku süresi ve stres seviyesi gibi metriklerin istatistiksel analizlerinde normal dağılım varsayımı önemli bir rol oynar. Q-Q plot, bu varsayımın doğruluğunu görsel olarak test ederek, hangi değişkenlerin dönüşüm veya alternatif analiz yöntemleri gerektirdiğini belirlemeyi sağlamış ve istatistiksel analiz süreçlerinin doğruluğunu artırmıştır.

Grafik, verinin teorik dağılımla ne kadar uyduğunu gösterir; veri noktaları, kırmızı referans çizgisine yakın bir dağılım gösteriyorsa, veri teorik dağılımla uyumludur. Eğer sapmalar gözlemlenirse, bu durum normal dağılım varsayımının ihlal edildiğine işaret etmektedir.



Şekil 14 Günlük Adım Miktarı QQ Plot

Şekil 14'teki incelendiğinde, günlük adım miktarı verisine ait Q-Q plot analizi sonuçları, verilerin normal dağılım varsayımıyla büyük ölçüde uyumlu olduğunu göstermektedir. Veri noktalarının çoğunluğunun referans çizgisi üzerinde veya yakınında seyretmesi, dağılımın normallik varsayımını desteklemektedir. Bu Q-Q plot analizi sonucunda, verilerin istatistiksel testler için gerekli olan normallik varsayımını karşıladığı ve ileri analizler için uygun olduğu değerlendirilmektedir.



Şekil 15 Yakılan Kalori Miktarı QQ Plot

Yakılan kalori verisine ait Q-Q plot analizi (Şekil 15), verilerin normal dağılımdan belirgin sapmalar gösterdiğini ortaya koymaktadır. Verinin dağılımı S-şeklinde bir örüntü sergilemektedir. Bu durum, dağılımın pozitif çarpıklık (sağa çarpıklık) gösterdiğini ve normal dağılım varsayımının ihlal edildiğini işaret etmektedir. Bu tür bir dağılım yapısı, testlerin uygulanmasında dikkatli olunması gerektiğini ve verilerin analizi için uygun veri dönüşümlerinin yapılması gerektiğini göstermektedir.

Tespit edilen normal dağılımdan sapmaların incelenmesi ve ileri istatistiksel analizlerin güvenilirliğini artırmak amacıyla uygulanması muhtemel veri dönüşüm yöntemleri, çalışmanın bir sonraki aşaması olan “normallik dönüşümleri” bölümünde detaylı olarak ele alınacaktır. Bu kapsamda gerçekleştirilecek dönüşümlerin etkinliği ve sonuçları, metodolojik bir yaklaşımla irdelenip yorumlanacaktır.

4.2 Normallik Testleri (Shapiro-Wilk ve Kolmogorov-Smirnov Testleri)

Verilerin normal dağılıma uygunluğunun analitik olarak değerlendirilmesi için Shapiro-Wilk ve Kolmogorov-Smirnov testleri kullanılmıştır. Shapiro-Wilk testi, örneklem büyüklüğünün 30’dan küçük olduğu durumlarda normallik varsayımını test etmede yüksek istatistiksel güce

sahipken, Kolmogorov-Smirnov testi özellikle büyük örneklerde tercih edilen bir yöntemdir. Bu çalışmada Kolmogorov-Smirnov testlerinden gelen sonuçlar dikkate alınacaktır. Her iki testte de sıfır hipotezi (H_0) verilerin normal dağılıma uyduğunu, alternatif hipotez (H_1) ise normal dağılımdan sapma olduğunu varsayar. Bu testlerde $p < 0.05$ değeri, verilerin normal dağılımdan anlamlı bir şekilde saptığını göstermektedir.

H_0 : Verilerin dağılımı normal dağılıma uymaktadır.
H_1 : Verilerin dağılımı normal dağılıma uymamaktadır.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
age	.072	3000	<.001	.953	3000	<.001
duration_minutes	.069	3000	<.001	.950	3000	<.001
calories_burned	.103	3000	<.001	.904	3000	<.001
avg_heart_rate	.049	3000	<.001	.987	3000	<.001
hours_sleep	.027	3000	<.001	.998	3000	.001
daily_steps	.023	3000	.001	.999	3000	.045
hydration_level	.084	3000	<.001	.951	3000	<.001
bmi	.038	3000	<.001	.989	3000	<.001
resting_heart_rate	.010	3000	.200*	1.000	3000	.823
blood_pressure_systolic	.017	3000	.034	.999	3000	.140
blood_pressure_diastolic	.014	3000	.200*	.999	3000	.016

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Şekil 16 Normallik Testi Sonuçları

Gerçekleştirilen normallik analizi sonuçları, veri setindeki değişkenlerin dağılım karakteristiklerini ortaya koymaktadır. Kolmogorov-Smirnov ve Shapiro-Wilk testlerinden elde edilen bulgular, çoğu değişkenin normal dağılım varsayımından anlamlı sapmalar gösterdiğini işaret etmektedir ($p < 0.001$). Yaş, aktivite süresi, yakılan kaloriler, ortalama kalp atış hızı, uyku süresi, hidrasyon seviyesi ve vücut kitle indeksi değişkenlerinin her iki test sonucunda da normal dağılımdan anlamlı derecede saptığı tespit edilmiştir. Bununla birlikte,

dinlenme kalp atış hızı (K-S $p=0.200$) ve diyastolik kan basıncı (K-S $p=0.200$) değişkenlerinin normal dağılım varsayımını karşıladığı görülmüştür. Bu bulgular ışığında, normal dağılım varsayımından sapma gösteren değişkenler için uygun dönüşüm tekniklerinin uygulanması, parametrik testlerin kullanılabilmesi açısından gerekli görülmektedir.

4.3 Normallik Dönüşümleri

Literatürde en çok kullanılan normallik dönüşüm yöntemlerinden bir tanesi **Rank-Based Normal Score Transformation**²'dir. Bu dönüşüm, verileri normal dağılıma mümkün olduğunca yakın bir forma dönüştürmek için sıralama (ranking) temelli bir yöntemdir. Bu yöntem, özellikle çarpık, uç değerlere (outliers) duyarlı veya kesikli verilerle çalışırken kullanılır. RBNST, verilerin sıralamasına dayalı olarak her bir gözlemi standart normal dağılımın z-skorlarına dönüştürür.

RBNST algoritmasının seçilmesinin temel nedenleri aşağıda verildiği gibidir:

1. Esneklik:
 - Hem kesikli hem de sürekli verilere uygulanabilir.
 - Veri tipleri ve dağılımları hakkında katı varsayımlar gerektirmez.
2. Uç Değerlerin Etkisini Azaltma:
 - Sıralama temelli olduğu için uç değerler (outliers) dönüşüm üzerinde minimum etki yapar.
3. Çarpıklığı Azaltma:
 - Sağ veya sol çarpık verilere sahip olduğunda, bu dönüşüm çarpıklığı büyük ölçüde azaltır.
4. Normal Dağılıma Yaklaşma:
 - Parametrik analizler (örneğin ANOVA, regresyon) için verilerin normal dağılıma uygunluğu önemlidir. RBNST, bu gereksinimi karşılamaya yardımcı olur.

² Chien, L. C. (2020). A rank-based normalization method with the fully adjusted full-stage procedure in genetic association studies. PloS one, 15(6), e0233847.

Veri setinde bulunan bütün sayısal değişkenlere RBNST yöntemi uygulanmış olup sonuçları Şekil 17’de verilmiştir.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
age_transformed	.016	3000	.058	.994	3000	<.001
duration_minutes_transformed	.009	3000	.200 [*]	.997	3000	<.001
calories_burned_transformed	.004	3000	.200 [*]	1.000	3000	1.000
avg_heart_rate_transformed	.013	3000	.200 [*]	1.000	3000	.910
hours_sleep_transformed	.021	3000	.003	.999	3000	.023
stress_level_transformed	.068	3000	<.001	.962	3000	<.001
daily_steps_transformed	.001	3000	.200 [*]	1.000	3000	1.000
hydration_level_transformed	.032	3000	<.001	.989	3000	<.001
bmi_transformed	.001	3000	.200 [*]	1.000	3000	1.000
resting_heart_rate_transformed	.006	3000	.200 [*]	1.000	3000	1.000
blood_pressure_systolic_transformed	.004	3000	.200 [*]	1.000	3000	1.000
blood_pressure_diastolic_transformed	.004	3000	.200 [*]	1.000	3000	1.000
fitness_level_transformed	.001	3000	.200 [*]	1.000	3000	1.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Şekil 17 Dönüşüm Sonrası Normallik Testi Sonuçları

Şekil 17’de sunulan normality testleri sonuçları, Rank-Based Normal Score Transformation (RBNST) yönteminin uygulanmasının ardından veri setindeki sayısal değişkenlerin dağılımlarına ilişkin değerlendirmeleri göstermektedir. Kolmogorov-Smirnov testine göre, p-değerleri incelendiğinde bazı değişkenlerin normal dağılıma yaklaştığı görülmektedir. Özellikle, calories_burned_transformed, avg_heart_rate_transformed, ve daily_steps_transformed gibi değişkenler normal dağılım varsayımını karşılamakta, ancak bazı değişkenlerde (stress_level_transformed ve hours_sleep_transformed) hala belirgin sapmalar bulunmaktadır.

RBNST yönteminin genel olarak verilerin dağılımını düzeltmek ve analitik yöntemlere uygun hale getirmek için etkili bir dönüşüm sağladığı, ancak bazı değişkenlerde tamamen normallik sağlanamadığı gözlenmiştir. Buna rağmen, dönüşümün veri seti genelindeki varyans sorunlarını ve çarpıklıkları azaltarak çok değişkenli analizlerin daha sağlam temellere dayandırılmasına olanak tanıdığı değerlendirilmektedir.

Bu noktadan itibaren, elde edilen dönüştürülmüş veri seti kullanılarak çalışmanın diğer aşamalarına devam edilecektir. Özellikle çok değişkenli istatistiksel analizler (örneğin MANOVA, faktör analizi, kümeleme) için dönüştürülmüş veri setinin uygun bir temel oluşturduğu düşünülmektedir. **Dönüşüm sonuçlarının yeterli düzeyde olduğu varsayımıyla (bütün değişkenlerin normal dağıldığı varsayımı altında)**, bundan sonraki analizlerde elde edilen veri seti kullanılacaktır.

5. MANOVA Analizi

5.1 Tek Yönlü MANOVA Analizi

Bu bölümde, veri setindeki bireylerin fitness seviyesi kategorilerine (fitness_level_category) göre belirli özellikler açısından anlamlı farklılık gösterip göstermediğini değerlendirmek amacıyla Tek Yönlü Çok Değişkenli Varyans Analizi (MANOVA) uygulanmıştır. MANOVA, birden fazla bağımlı değişken arasında ilişkiler göz önüne alınarak, gruplar arasında eşzamanlı karşılaştırma yapılmasını sağlayan güçlü bir yöntemdir. Bu analizde, bağımlı değişkenler olarak aktivitenin yapılma süresi (duration_minutes_transformed) ve günlük atılan adım miktarı (daily_steps_transformed) kullanılmıştır.

Analizin amacı, bireylerin fitness seviyesi kategorilerinin bu iki bağımlı değişken üzerinde anlamlı bir etkisinin olup olmadığını değerlendirmektir. MANOVA, her bir bağımlı değişken için ayrı ayrı varyans analizine göre daha duyarlıdır ve bağımlı değişkenler arasındaki potansiyel ilişkiyi de hesaba katar. Bu nedenle, fitness seviyesi kategorilerinin, aktivitenin yapılma süresi ve atılan adım sayısı gibi fiziksel parametreler üzerindeki etkisini daha kapsamlı bir şekilde incelemek için uygun bir yöntemdir. Analiz sonuçları, fitness seviyesinin bireylerin fiziksel özellikleri üzerindeki etkisini daha iyi anlamak ve bu bilgileri ilgili stratejik kararlar veya müdahaleler için kullanmak adına önemli ipuçları sağlayacaktır.

Between-Subjects Factors		
		N
fitness_level_category	High	999
	Low	1000
	Medium	1001

Şekil 18 Kategorik Değişkenin Dağılımı

Fitness seviyesi kategorilerinin dağılımını incelendiğinde, örneklemdeki katılımcıların dengeli bir şekilde dağıldığı gözlemlenmiştir. Yüksek fitness seviyesinde 999, düşük fitness seviyesinde 1000 ve orta fitness seviyesinde 1001 katılımcı bulunmaktadır. Bu dağılım, her bir kategori için neredeyse eşit sayıda gözlem içermekte olup, bu durum gruplar arası karşılaştırmaların istatistiksel güvenilirliğini artırmakta ve potansiyel yanlılıkları minimize etmektedir.

Descriptive Statistics				
	fitness_level_category	Mean	Std. Deviation	N
duration_minutes_transformed	High	.03299326474	1.0003930468	999
	Low	.02001463348	.98562157081	1000
	Medium	-.0533324225	.99076272957	1001
	Total	-.0001369500	.99267223171	3000
daily_steps_transformed	High	-.6915435226	.84110778680	999
	Low	.42319002563	.85816240316	1000
	Medium	.26739572907	.91022590235	1001
	Total	.00000039045	.99994678595	3000

Şekil 19 Tanımlayıcı İstatistikler

**Box's Test of
Equality of
Covariance
Matrices^a**

Box's M	10.373
F	1.727
df1	6
df2	223855470.80
Sig.	.110

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + fitness_level_category

Şekil 20 Kovaryans Matrislerinin Eşitliği Testi

H_0 : Gruplar arası varyans – kovaryans matrisi birbirine eşittir.

H_1 : Gruplar arası varyans – kovaryans matrisi birbirinden farklıdır.

Box's M testi sonuçlarını değerlendirdiğimizde, test istatistiği değeri 10.373 ve buna karşılık gelen F değeri 1.727 olarak hesaplanmıştır. Test sonucunda elde edilen anlamlılık düzeyi $p=0.110$ ($p>0.05$) olarak bulunmuştur. Bu sonuç, kovaryans matrislerinin gruplar arasında eşit olduğunu göstermektedir, yani sıfır hipotezi (H_0) reddedilememiştir.

Bu durumda, çok deęişkenli analizlerde Wilks' Lambda test istatistięinin kullanılması uygundur, çünkü kovaryans matrislerinin eşitlięi varsayımı saęlanmışır. Analizlere bu doęrultuda devam edilmiştir.

Multivariate Tests ^a							
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.000	.000 ^b	2.000	2996.000	1.000	.000
	Wilks' Lambda	1.000	.000 ^b	2.000	2996.000	1.000	.000
	Hotelling's Trace	.000	.000 ^b	2.000	2996.000	1.000	.000
	Roy's Largest Root	.000	.000 ^b	2.000	2996.000	1.000	.000
fitness_level_category	Pillai's Trace	.244	208.358	4.000	5994.000	<.001	.122
	Wilks' Lambda	.756	224.725 ^b	4.000	5992.000	<.001	.130
	Hotelling's Trace	.322	241.239	4.000	5990.000	<.001	.139
	Roy's Largest Root	.321	481.169 ^c	2.000	2997.000	<.001	.243

a. Design: Intercept + fitness_level_category

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Şekil 21 Çok Deęişkenli Test Sonuçları

H_0 : Günlük adım sayısı ve aktivite süresi açısından, farklı fitness seviyeleri kategorileri arasında istatistiksel olarak anlamlı bir fark yoktur.

H_1 : Günlük adım sayısı ve (veya) aktivite süresi açısından, en az bir fitness seviyeleri kategorisi diğerlerinden istatistiksel olarak anlamlı bir fark göstermektedir.

Çok deęişkenli test sonuçları tablosu (Şekil 21) incelendięinde, fitness seviyesi kategorisinin etkisini deęerlendiren dört farklı test istatistięi (Pillai's Trace, Wilks' Lambda, Hotelling's Trace ve Roy's Largest Root) de istatistiksel olarak anlamlı bulunmuştur, yani H_0 hipotezi reddedilmiştir. ($p<0.001$).

Wilks' Lambda deęeri 0.756 olarak hesaplanmış ve $F(4,5992)=224.725$ deęeri elde edilmiştir. Kısmi eta kare deęeri 0.130 olarak bulunmuş olup, bu deęer orta düzeyde bir etki büyüklüęüne işaret etmektedir. Baęımlı deęişkenlerin %13'ü grup deęişkeni tarafından açıklanabilir.

Bu sonuçlar, fitness seviyesi kategorilerinin bağımlı değişkenler üzerinde anlamlı bir etkiye sahip olduğunu göstermektedir.

Levene's Test of Equality of Error Variances ^a					
		Levene Statistic	df1	df2	Sig.
duration_minutes_transformed	Based on Mean	.522	2	2997	.593
	Based on Median	.523	2	2997	.593
	Based on Median and with adjusted df	.523	2	2996.201	.593
	Based on trimmed mean	.522	2	2997	.593
daily_steps_transformed	Based on Mean	2.575	2	2997	.076
	Based on Median	2.591	2	2997	.075
	Based on Median and with adjusted df	2.591	2	2979.179	.075
	Based on trimmed mean	2.576	2	2997	.076

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + fitness_level_category

Şekil 22 Levene Testi Sonuçları

Varyans homojenliğini değerlendirmek amacıyla Şekil 22'deki Levene testi kullanılmıştır. Bu testin hipotezleri;

H_0 : Gruplar arasındaki hata varyansları eşittir.
H_1 : Gruplar arasındaki hata varyansları eşit değildir.

şeklindedir. Levene testi sonuçlarına göre, hem duration_minutes_transformed hem de daily_steps_transformed değişkenleri için anlamlılık düzeyleri (p-değeri) 0.05'in üzerinde bulunmuştur. Bu sonuç, varyansların gruplar arasında eşit olduğunu gösterir ve dolayısıyla H_0 reddedilemez. Varyansların homojen olduğu sonucuna varıldığından, gruplar arasındaki farkları belirlemek için post-hoc analizde **Tukey testi** uygulanması uygun görülmüştür.

Tests of Between-Subjects Effects							
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	duration_minutes_transformed	4.335 ^a	2	2.168	2.201	.111	.001
	daily_steps_transformed	728.416 ^b	2	364.208	480.795	<.001	.243
Intercept	duration_minutes_transformed	3.511E-5	1	3.511E-5	.000	.995	.000
	daily_steps_transformed	.000	1	.000	.000	.984	.000
fitness_level_category	duration_minutes_transformed	4.335	2	2.168	2.201	.111	.001
	daily_steps_transformed	728.416	2	364.208	480.795	<.001	.243
Error	duration_minutes_transformed	2950.874	2997	.985			
	daily_steps_transformed	2270.265	2997	.758			
Total	duration_minutes_transformed	2955.209	3000				
	daily_steps_transformed	2998.681	3000				
Corrected Total	duration_minutes_transformed	2955.209	2999				
	daily_steps_transformed	2998.681	2999				

a. R Squared = .001 (Adjusted R Squared = .001)

b. R Squared = .243 (Adjusted R Squared = .242)

Şekil 23 Between-Subjects Effects Test Sonuçları

Bağımsız gruplar arasındaki etkileri değerlendirmek için yapılan 'Between-Subjects Effects' testi sonuçlarına göre (Şekil 23), hipotezler şu şekilde belirlenmiştir:

H_0 : Fitness seviyeleri kategorileri arasında bağımlı değişkenlerde anlamlı bir fark yoktur.

H_1 : Fitness seviyeleri kategorileri arasında bağımlı değişkenlerde en az bir bağımlı değişken için anlamlı bir fark vardır.

Aktivite süresi değişkeni için fitness seviyeleri kategorisinin etkisi istatistiksel olarak anlamlı bulunmamıştır ($F(2, 2997) = 2.201, p = 0.111$). Bu sonuç, Sig. değeri 0.05'ten büyük olduğu için H_0 reddedilemez ve gruplar arasında aktivite süresi açısından anlamlı bir fark olmadığı anlamına gelir. Buna karşılık, günlük adım sayısı değişkeni için fitness seviyeleri kategorisi üzerinde anlamlı bir etki tespit edilmiştir ($F(2, 2997) = 480.795, p < 0.001$). Sig. değeri 0.05'ten küçük olduğu için bu durumda H_0 reddedilir ve gruplar arasında günlük adım sayısı açısından anlamlı farklılıklar olduğu sonucuna varılır. Ek olarak, Partial Eta Squared değeri, fitness seviyeleri kategorisinin günlük adım sayısı üzerindeki varyansın %24.3'ünü açıkladığını

göstermektedir. Bu bulgu, fitness seviyelerinin günlük adım sayısı üzerindeki etkisinin oldukça güçlü olduğunu işaret etmektedir.

Post Hoc Tests

fitness_level_category

Multiple Comparisons							
Tukey HSD							
Dependent Variable	(I) fitness_level_category	(J) fitness_level_category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
duration_minutes_transfor med	High	Low	.01297863126	.04438698092	.954	-.0911029763	.11706023884
		Medium	.08632568719	.04437589942	.126	-.0177299357	.19038131013
	Low	High	-.0129786313	.04438698092	.954	-.1170602388	.09110297633
		Medium	.07334705594	.04436479296	.224	-.0306825238	.17737663569
	Medium	High	-.0863256872	.04437589942	.126	-.1903813101	.01772993574
		Low	-.0733470559	.04436479296	.224	-.1773766357	.03068252382
daily_steps_transformed	High	Low	-1.114733548*	.03893306040	<.001	-1.206026433	-1.023440664
		Medium	-.9589392517*	.03892334051	<.001	-1.050209344	-.8676691592
	Low	High	1.114733548*	.03893306040	<.001	1.0234406639	1.2060264326
		Medium	.1557942966*	.03891359873	<.001	.06454704729	.24704154584
	Medium	High	.9589392517*	.03892334051	<.001	.86766915919	1.0502093441
		Low	-1.1557942966*	.03891359873	<.001	-.2470415458	-.0645470473
Based on observed means. The error term is Mean Square(Error) = .758. *. The mean difference is significant at the .05 level.							

Şekil 24 Tukey HSD Testi Sonuçları

Tukey HSD testi sonuçları (Şekil 24), farklı fitness seviyeleri kategorileri arasında yapılan çoklu karşılaştırmaların sonuçlarını göstermektedir.

H_0 : Gruplar arasında anlamlı bir fark yoktur.
H_1 : Gruplar arasında anlamlı bir fark vardır.

Aktivite süresi değişkeni için tüm grup karşılaştırmalarının anlamlılık düzeyleri (p-değeri) 0.05'in üzerinde bulunmuştur. Bu sonuçlar, yüksek, düşük ve orta fitness seviyeleri kategorileri arasında aktivite süresi açısından istatistiksel olarak anlamlı bir fark olmadığını göstermektedir. Diğer bir deyişle, farklı fitness seviyeleri bu değişken üzerinde belirgin bir farklılık yaratmamaktadır; bu nedenle H_0 reddedilemez.

Öte yandan, günlük adım sayısı değişkeni için, fitness seviyeleri kategorileri arasında istatistiksel olarak anlamlı farklar tespit edilmiştir. Özellikle, yüksek fitness seviyesine sahip bireyler ile düşük ve orta seviyedeki bireyler arasında ($p < 0.001$) ve düşük seviyedeki bireyler

ile orta seviyedeki bireyler arasında ($p < 0.001$) anlamlı farklar bulunmuştur. Bu bulgular, farklı fitness seviyelerinin günlük adım sayısı üzerinde önemli bir etki yarattığını göstermektedir. Sonuç olarak, günlük adım sayısı açısından fitness seviyeleri kategorileri arasında anlamlı farklılıklar gözlemlenmiştir ve H_0 reddedilir.

Homogeneous Subsets

duration_minutes_transformed

Tukey HSD^{a,b,c}

fitness_level_category	N	Subset 1
Medium	1001	-.0533324225
Low	1000	.02001463348
High	999	.03299326474
Sig.		.126

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = .985.
a. Uses Harmonic Mean Sample Size = 999.999.
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
c. Alpha = .05.

Şekil 25 Homeojen Alt Örnekler (Aktivite Süresi)

Tukey HSD testi sonuçları, aktivite süresi değişkeni için farklı fitness seviyeleri kategorilerinin ortalama değerlerini göstermektedir. Her bir kategoriye ait gözlem sayıları şu şekildedir: Orta (Medium) kategori için 1001, Düşük (Low) kategori için 1000, ve Yüksek (High) kategori için 999 gözlem. Alt kümeler (Subsets) içindeki ortalama değerler, fitness seviyelerinin bu değişken üzerindeki etkisini ifade etmektedir.

H_0 : Farklı fitness seviyeleri kategorileri (High, Medium, Low) arasında aktivite süresi açısından anlamlı bir fark yoktur.

H_1 : Farklı fitness seviyeleri kategorileri (High, Medium, Low) arasında aktivite süresi açısından en az bir çift kategori arasında anlamlı bir fark vardır.

Testin anlamlılık düzeyi (Sig.) 0.126 olarak bulunmuştur. Bu değer 0.05'ten büyük olduğundan, gruplar arasında aktivite süresi açısından istatistiksel olarak anlamlı bir fark olmadığı sonucuna varılır ve bu nedenle H_0 hipotezi reddedilemez: Gruplar arasında anlamlı bir fark yoktur. Bu sonuç, üç fitness seviyesi kategorisinin, ortalama aktivite değerleri açısından birbirine benzer olduğunu göstermektedir.

Sonuç olarak, fitness seviyelerinin aktivite süresi değişkeni üzerindeki etkisi anlamlı bulunmamış ve gruplar arasında fark olmadığı belirlenmiştir. Ancak, bu analiz sadece belirtilen değişkenle sınırlıdır ve diğer potansiyel faktörler de dikkate alınarak daha derinlemesine analizler yapılması önerilir.

daily_steps_transformed

Tukey HSD^{a,b,c}

fitness_level_category	N	1	2	3
High	999	-.6915435226		
Medium	1001		.26739572907	
Low	1000			.42319002563
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = .758.
a. Uses Harmonic Mean Sample Size = 999.999.
b. The group sizes are unequal. The harmonic mean of the group sizes is used.
Type I error levels are not guaranteed.
c. Alpha = .05.

Şekil 26 Homeojen Alt Örnekler (Günlük Adım Miktarı)

Şekil 26'da günlük adım miktarı değişkeni için farklı fitness seviyeleri kategorileri (High, Medium, Low) arasındaki karşılaştırmaları göstermektedir.

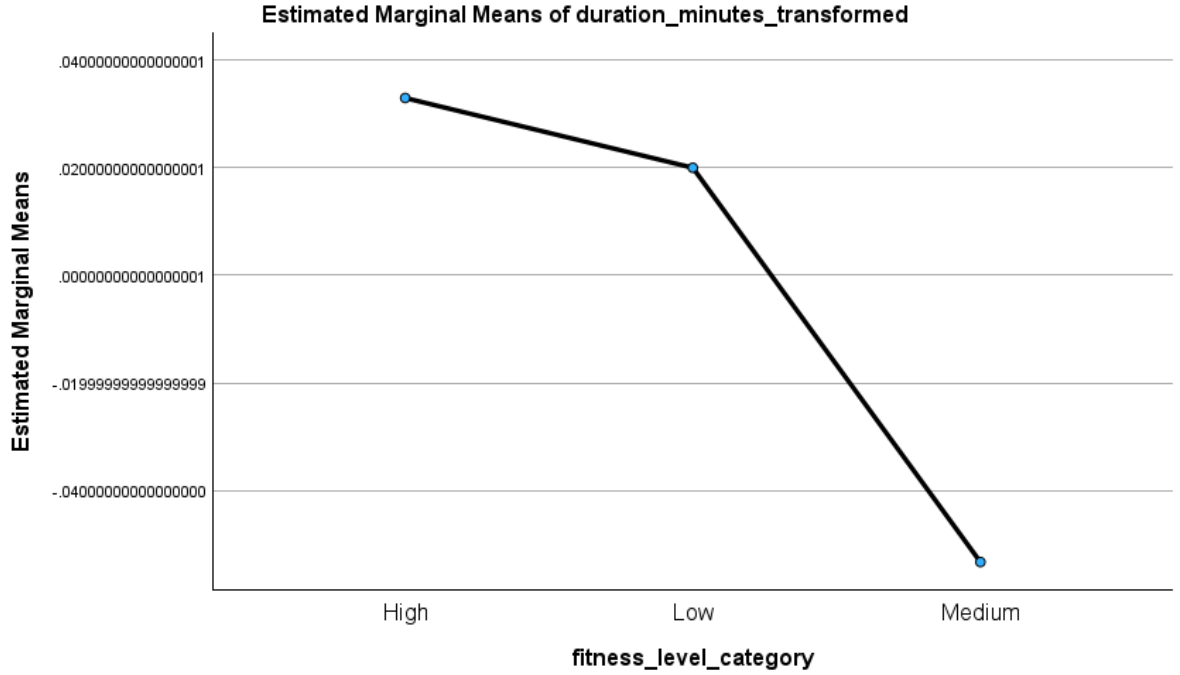
H_0 : Farklı fitness seviyeleri kategorileri (High, Medium, Low) arasında günlük adım miktarı bakımından anlamlı bir fark yoktur.

H_1 : Farklı fitness seviyeleri kategorileri (High, Medium, Low) arasında günlük adım miktarı bakımından en az bir çift kategori arasında anlamlı bir fark vardır.

Tukey HSD testinin sonuçlarına göre (Şekil 26), her bir fitness seviyesi kategorisi için alt kümeler (subsets) belirtilmiştir. Testin anlamlılık düzeyi (Sig.) 1.000 olarak bulunmuştur, bu da tüm gruplar arasında anlamlı bir fark olmadığını göstermektedir. Özellikle, alt kümeler arasında herhangi bir çift kategori için p-değerinin 0.05'in altında olmaması, H_0 hipotezinin reddedilemeyeceğini ve gruplar arasında günlük adım miktarı açısından istatistiksel olarak anlamlı bir fark olmadığını gösterir.

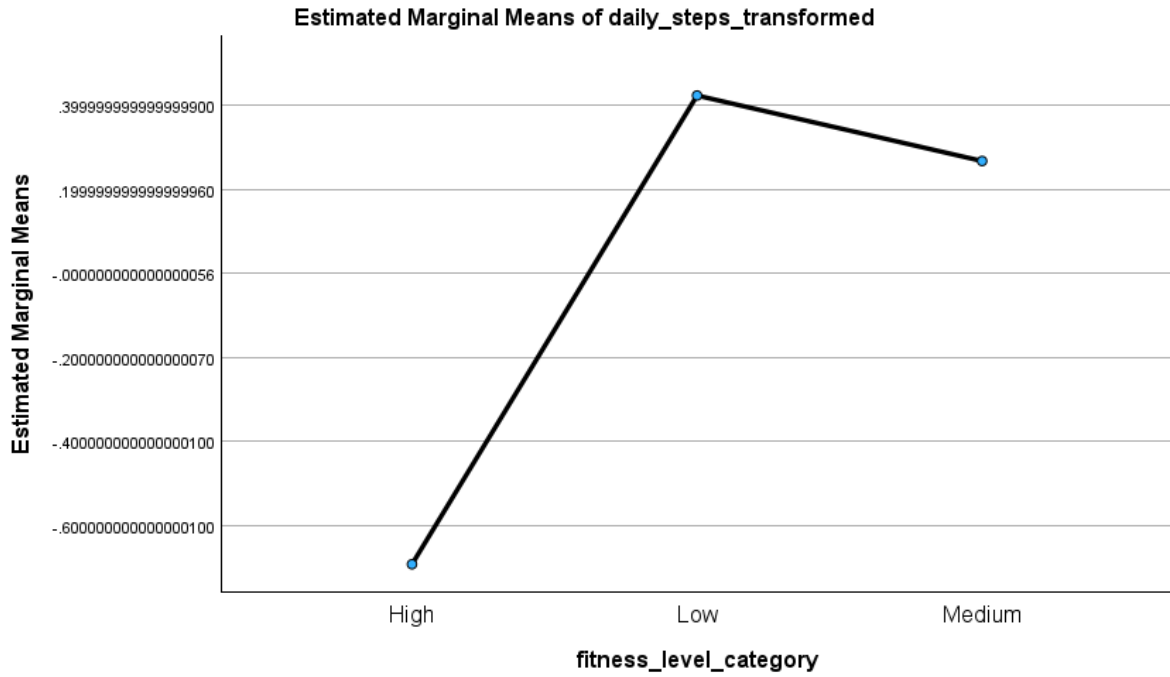
Profile Plots

duration_minutes_transformed



Şekil 27 Tahmini Marjin Ortalamaları (Aktivite Süresi)

Grafik 27, aktivite süresi değişkeninin tahmini marjinal ortalamalarını farklı fitness seviyesi kategorilerine (High, Low, Medium) göre göstermektedir. Grafikte, High fitness seviyesindeki bireylerin tahmini marjinal ortalamasının en yüksek olduğu, bunu Low ve Medium fitness seviyelerinin takip ettiği açıkça görülmektedir. Bu durum, fitness seviyeleri ile süre arasındaki ilişkiyi gözler önüne sermektedir.



Şekil 28 Tahmini Marjin Ortalamaları (Günlük Adım Sayısı)

Tahmini Marjin Ortalamaları (Günlük Adım Sayısı) grafiği incelendiğinde (Şekil 28), farklı fitness seviyesi kategorilerinin ("High", "Low", "Medium") günlük adım sayısının değerlerine göre tahmini marjinal ortalamalarını göstermektedir. Grafikte, "Low" fitness seviyesindeki bireylerin tahmini marjinal ortalamasının en yüksek olduğu, bunu sırasıyla "Medium" ve "High" fitness seviyelerinin takip ettiği gözlemlenmektedir. Bu eğilim, fitness seviyeleri ile günlük adım sayısı arasında ters bir ilişki olabileceğini düşündürmektedir.

5.2 Çift Yönlü MANOVA Analizi

Bireylerin fiziksel aktivite davranışlarını anlamak amacıyla, çift yönlü MANOVA yöntemi uygulanmıştır. Analiz, bağımsız değişkenler olarak fitness seviyesi kategorisi (fitness_level_category) ve aktivite türü (activity_type) ile bağımlı değişkenler olan aktivite süresi (duration_minutes_transformed) ve günlük adım sayısı (daily_steps_transformed) arasındaki ilişkileri incelemeyi hedeflemektedir. Bu yöntem, bağımlı değişkenler arasındaki potansiyel korelasyonu dikkate alarak, bağımsız değişkenlerin her iki bağımlı değişken üzerindeki bireysel ve tümleşik etkilerini kapsamlı bir şekilde değerlendirme imkânı sunmaktadır. Analizin temel amacı, fitness seviyelerinin ve farklı aktivite türlerinin bağımlı değişkenler üzerindeki anlamlı etkilerini belirlemek ve bu değişkenlerin etkileşimlerinin fiziksel aktivite davranışlarını nasıl şekillendirdiğini ortaya koymaktır.

Box's Test of Equality of Covariance Matrices^a

Box's M	96.541
F	1.101
df1	87
df2	6330437.154
Sig.	.243

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept
+
fitness_level_cat
egory +
activity_type +
fitness_level_cat
egory *
activity_type

Şekil 29 Kovaryans Matrislerinin Eşitliği Testi

Box's Test of Equality of Covariance Matrices, çok deęişkenli varyans analizinin (MANOVA) temel varsayımlarından biri olan kovaryans matrislerinin eşitlięi varsayımını test etmek için kullanılan bir istatistiksel yöntemdir.

*H_0 : Baęımlı deęişkenler için gözlemlenen kovaryans matrisleri,
tüm gruplar arasında eşittir.*

*H_1 : Baęımlı deęişkenler için gözlemlenen kovaryans matrisleri,
gruplar arasında eşit deęildir.*

Yukarıdaki tabloda, Box's M testi sonucu 96.541, $F = 1.101$, ve anlamlılık düzeyi (Sig.) 0.243 olarak belirtilmiştir. p-deęeri (Sig. = 0.243), 0.05 anlamlılık düzeyinden büyük olduğundan, H_0 hipotezi reddedilemez. Bu durum, baęımlı deęişkenlerin kovaryans matrislerinin gruplar arasında eşit olduğu varsayımını desteklemektedir.

Sonuç olarak MANOVA analizi için kovaryans matrislerinin eşitlięi varsayımının sağlandığını görülmekte ve analiz sonuçlarının geçerlilięi açısından uygun bir temel oluşturmaktadır. Dolayısıyla, baęımsız deęişkenlerin baęımlı deęişkenler üzerindeki etkilerini deęerlendirmek üzere MANOVA'nın uygulanması istatistiksel olarak uygundur.

Multivariate Tests ^a							
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.000	.008 ^b	2.000	2969.000	.992	.000
	Wilks' Lambda	1.000	.008 ^b	2.000	2969.000	.992	.000
	Hotelling's Trace	.000	.008 ^b	2.000	2969.000	.992	.000
	Roy's Largest Root	.000	.008 ^b	2.000	2969.000	.992	.000
fitness_level_category	Pillai's Trace	.246	208.137	4.000	5940.000	<.001	.123
	Wilks' Lambda	.754	224.648 ^b	4.000	5938.000	<.001	.131
	Hotelling's Trace	.325	241.310	4.000	5936.000	<.001	.140
	Roy's Largest Root	.324	481.378 ^c	2.000	2970.000	<.001	.245
activity_type	Pillai's Trace	.007	1.138	18.000	5940.000	.307	.003
	Wilks' Lambda	.993	1.138 ^b	18.000	5938.000	.307	.003
	Hotelling's Trace	.007	1.138	18.000	5936.000	.307	.003
	Roy's Largest Root	.005	1.613 ^c	9.000	2970.000	.106	.005
fitness_level_category * activity_type	Pillai's Trace	.013	1.092	36.000	5940.000	.325	.007
	Wilks' Lambda	.987	1.092 ^b	36.000	5938.000	.325	.007
	Hotelling's Trace	.013	1.092	36.000	5936.000	.325	.007
	Roy's Largest Root	.009	1.423 ^c	18.000	2970.000	.110	.009

a. Design: Intercept + fitness_level_category + activity_type + fitness_level_category * activity_type

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Şekil 30 Çok Değişkenli Test Sonuçları

Tablodaki Multivariate Tests sonuçlarına göre (Şekil 30), bağımsız değişkenlerin etkileşimlerinin (fitness_level_category * activity_type) bağımlı değişkenler üzerinde anlamlı bir etkisi olup olmadığını değerlendirmek mümkündür.

H_0 : Fitness seviyesi kategorisi ile aktivite türü etkileşiminin bağımlı değişkenler üzerinde anlamlı bir etkisi yoktur.

H_1 : Fitness seviyesi kategorisi ile aktivite türü etkileşimi, bağımlı değişkenlerden en az biri üzerinde anlamlı bir etkiye sahiptir

Fitness seviyesi ve aktivite tipi etkileşimine ilişkin Wilks Lambda anlamlılık düzeyi (Sig.=0.325), bu etkileşimin bağımlı değişkenler üzerinde anlamlı bir etkisinin olmadığını ortaya koymaktadır. Yani H_0 hipotezi reddedilememiştir. Bu durumda, fitness seviyesi ve aktivite türü arasındaki etkileşim bağımlı değişkenler açısından anlamlı bir farklılık yaratmamaktadır. Partial Eta Squared değeri 0.007 olarak hesaplanmış olup, bu etkileşimin bağımlı değişkenlerin toplam varyansını açıklamadaki etkisinin %0.7 gibi düşük bir oranda olduğunu göstermektedir.

Levene's Test of Equality of Error Variances^a

		Levene Statistic	df1	df2	Sig.
duration_minutes_transformed	Based on Mean	.659	29	2970	.918
	Based on Median	.658	29	2970	.918
	Based on Median and with adjusted df	.658	29	2921.912	.918
	Based on trimmed mean	.665	29	2970	.913
daily_steps_transformed	Based on Mean	1.317	29	2970	.120
	Based on Median	1.277	29	2970	.147
	Based on Median and with adjusted df	1.277	29	2888.670	.147
	Based on trimmed mean	1.315	29	2970	.121

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + fitness_level_category + activity_type + fitness_level_category * activity_type

Şekil 31 Levene Testi Sonuçları

Tablo 31’de gösterilen Levene's Test of Equality of Error Variances tablosunda, bağımlı değişkenlerin gruplar arası hata varyanslarının eşit olup olmadığını test edilmektedir.

H_0 : Gruplar arasında hata varyansları eşittir.
H_1 : Gruplar arasında hata varyansları eşit değildir.

Bu hipotezler doğrultusunda, tablodaki anlamlılık (Sig.) değerleri incelenmiştir. Aktivite süresi değişkeni için tüm test yöntemlerine (Based on Mean, Median, Trimmed Mean) ait anlamlılık değerlerinin 0.05’ten büyük olduğu görülmektedir. Bu nedenle, H_0 hipotezi reddedilemez ve gruplar arasında hata varyanslarının eşit olduğu sonucuna varılmıştır. Benzer şekilde, günlük adım sayısı değişkeni için de tüm test yöntemlerine ait anlamlılık değerleri 0.05’ten büyük olduğundan H_0 hipotezinin reddedilemeyeceğini göstermektedir.

Sonuç olarak, her iki bağımlı değişken açısından da gruplar arasında hata varyanslarının eşit olduğu görülmektedir. Bu bulgu, varyansların homojenliğini varsayan post-hoc yöntemlerin kullanılabileceğini desteklemektedir. Tukey yöntemi, hata varyanslarının eşit olduğu durumlarda tercih edilen bir yöntem olup, bu analizde güvenle kullanılabilir.

Tests of Between-Subjects Effects							
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	duration_minutes_transformed	27.501 ^a	29	.948	.962	.524	.009
	daily_steps_transformed	755.729 ^b	29	26.060	34.507	<.001	.252
Intercept	duration_minutes_transformed	.002	1	.002	.002	.962	.000
	daily_steps_transformed	.010	1	.010	.013	.908	.000
fitness_level_category	duration_minutes_transformed	4.294	2	2.147	2.178	.113	.001
	daily_steps_transformed	726.404	2	363.202	480.933	<.001	.245
activity_type	duration_minutes_transformed	9.561	9	1.062	1.078	.376	.003
	daily_steps_transformed	8.187	9	.910	1.205	.287	.004
fitness_level_category * activity_type	duration_minutes_transformed	13.513	18	.751	.762	.748	.005
	daily_steps_transformed	19.346	18	1.075	1.423	.110	.009
Error	duration_minutes_transformed	2927.708	2970	.986			
	daily_steps_transformed	2242.952	2970	.755			
Total	duration_minutes_transformed	2955.209	3000				
	daily_steps_transformed	2998.681	3000				
Corrected Total	duration_minutes_transformed	2955.209	2999				
	daily_steps_transformed	2998.681	2999				

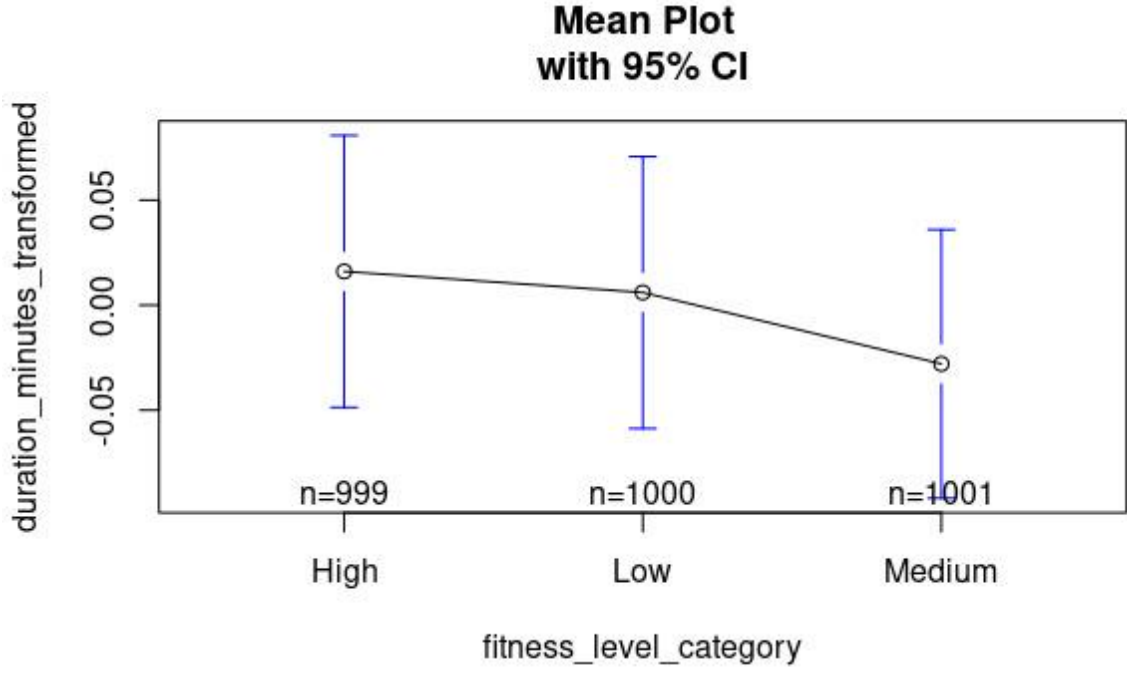
a. R Squared = .009 (Adjusted R Squared = .000)
b. R Squared = .252 (Adjusted R Squared = .245)

Şekil 32 Between-Subjects Effects Test Sonuçları

Denekler arası etki testleri kapsamında gerçekleştirilen analiz sonuçları değerlendirildiğinde (Şekil 32),

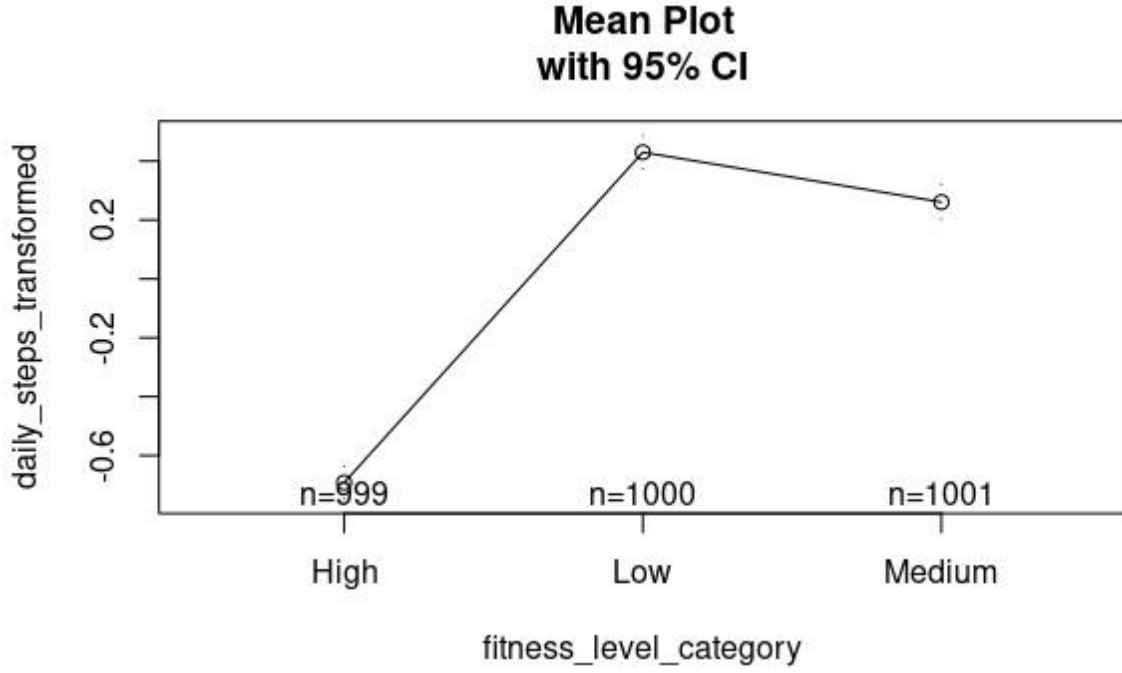
H_0 : Gruplar arasında anlamlı bir etki yoktur.
H_1 : Gruplar arasında anlamlı bir etki vardır.

Analiz sonuçları üzerinden, günlük adım sayısı değişkeni için fitness seviyesi kategorileri arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir ($p < 0.001$). Öte yandan, aktivite süresi değişkeni için gruplar arasında istatistiksel olarak anlamlı bir fark bulunmamıştır ($p = 0.113$). Aktivite türü ve fitness seviyesi kategorisi arasındaki etkileşim etkisi de istatistiksel olarak anlamlı bulunmamıştır ($p = 0.748$ ve $p = 0.110$, sırasıyla). Bu bulgular, fitness seviyesi kategorilerinin özellikle günlük adım sayısı üzerinde belirgin bir etkiye sahip olduğunu, ancak aktivite süresi üzerindeki etkisinin sınırlı olduğunu göstermektedir.



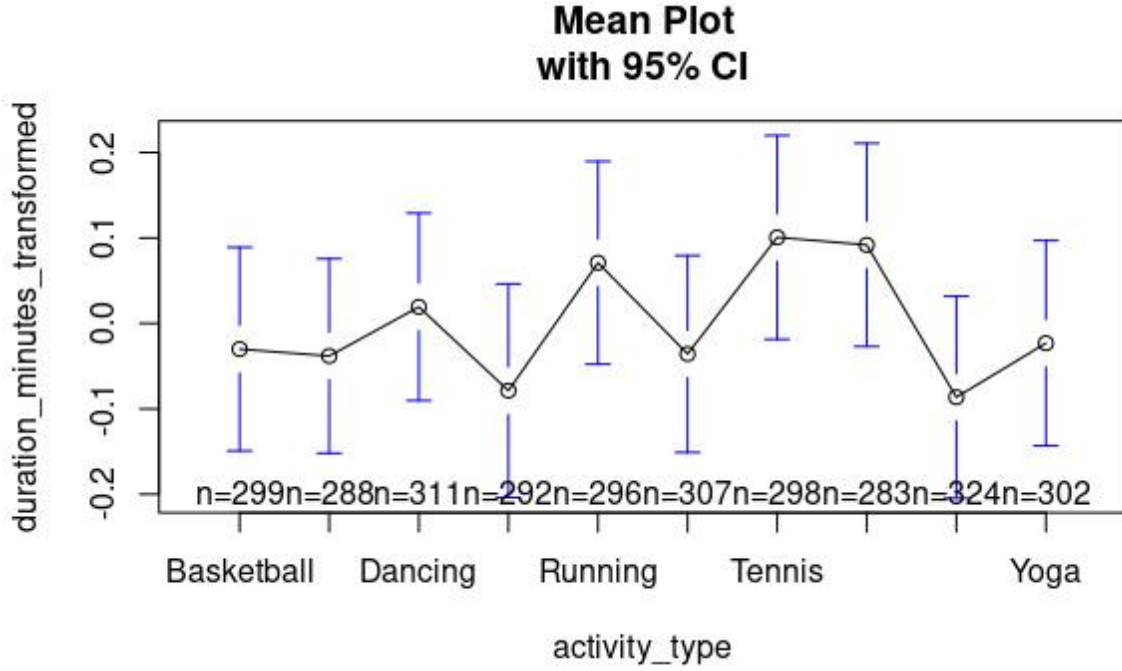
Şekil 33 Aktivite Süresi Değişkeninin Fitness Seviyesi Kategorilerine göre Dağılımı

Aktivite süresi değişkeninin fitness seviyesi kategorilerine göre dağılımını gösteren ortalama grafiği incelendiğinde (Şekil 33), yüksek fitness seviyesine sahip bireylerin aktivite süresi ortalamasının en yüksek (yaklaşık 0.01), orta fitness seviyesine sahip bireylerin ise en düşük (yaklaşık -0.03) değere sahip olduğu gözlemlenmiştir. Fitness seviyesi düştükçe aktivite süresinin de azalma eğiliminde olduğu, ancak bu ilişkinin güven aralıklarının örtüşmesi nedeniyle istatistiksel olarak anlamlı olmayabileceği görülmektedir. Bu bulgu, daha önceki istatistiksel analizlerde elde edilen sonuçları görsel olarak desteklemektedir.



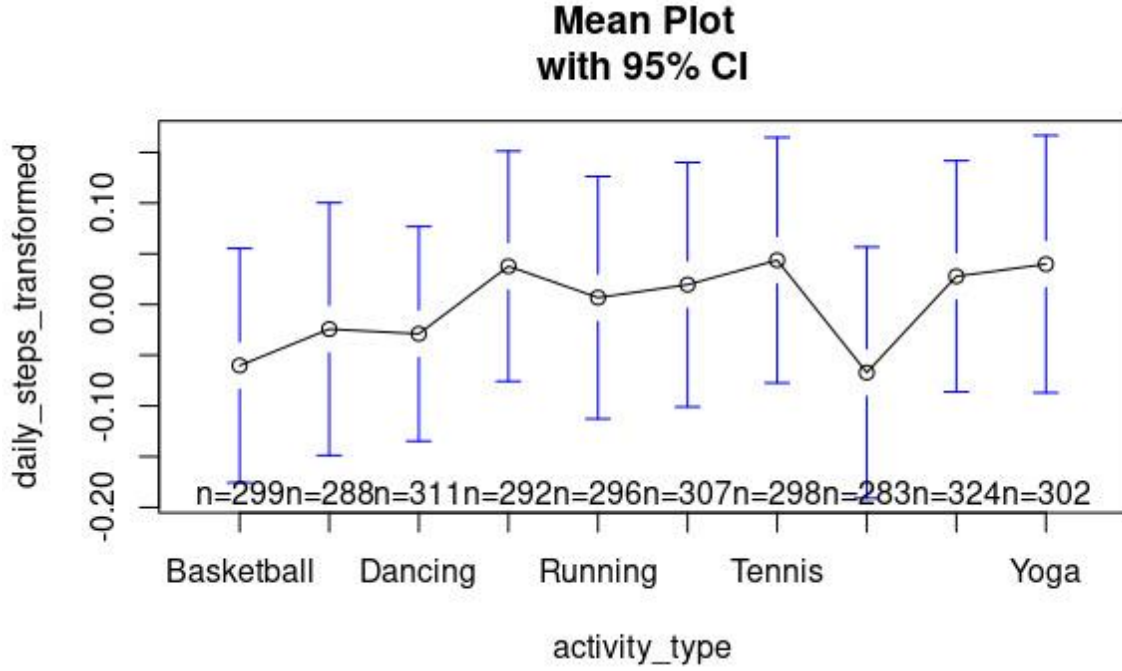
Şekil 34 Günlük Adım Sayısı Değişkeninin Fitness Seviyesi Kategorilerine göre Dağılımı

Şekil 34 incelendiğinde, günlük adım sayısı değişkeninin fitness seviyesi kategorilerine göre dağılımını gösteren ortalama grafikte fitness seviyesi yüksek olan bireylerde günlük adım sayısı ortalamasının en düşük (yaklaşık -0.6), düşük fitness seviyesine sahip bireylerde ise en yüksek (yaklaşık 0.3) değere sahip olduğu gözlemlenmiştir. Gruplar arasındaki farklılıkların belirgin olduğu ve istatistiksel olarak anlamlı olduğu görülmektedir. Bu durum, fitness seviyesi ile günlük adım sayısı arasında negatif yönlü bir ilişki olduğunu göstermektedir. Elde edilen bu görsel bulgu, önceki istatistiksel analizlerde tespit edilen anlamlı grup farklarını destekler niteliktedir.



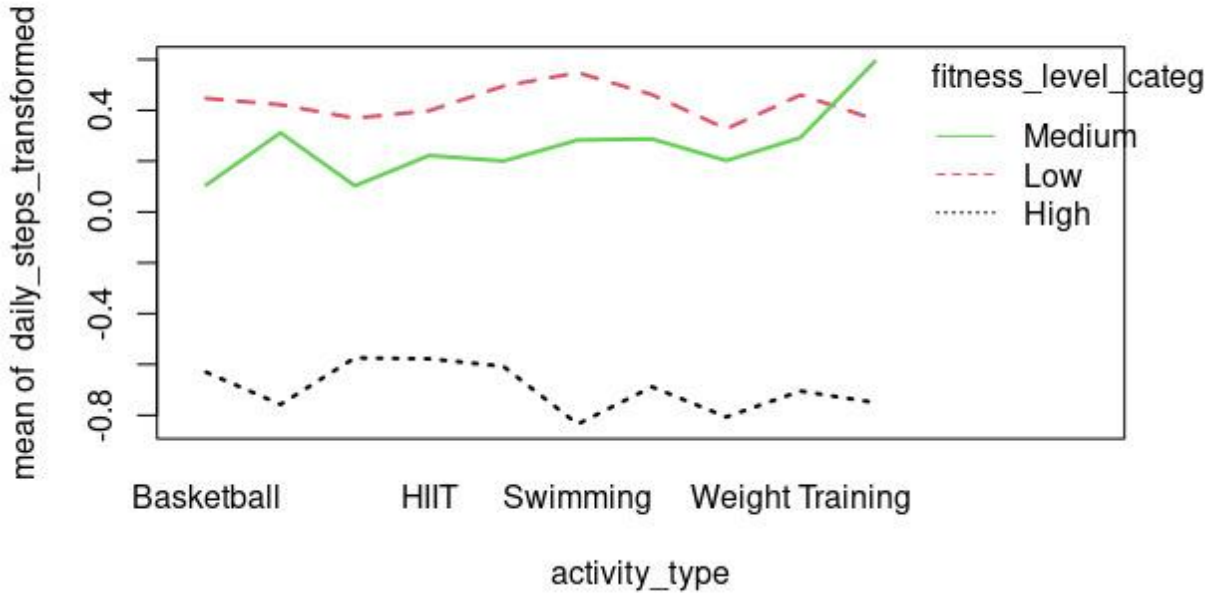
Şekil 35 Farklı Aktivite Türlerinin Aktivite Süresi Üzerindeki Etkisi

Farklı aktivite türlerinin aktivite süresi üzerindeki etkisini gösteren ortalama grafiği incelendiğinde (Şekil 35), aktivite türleri arasında aktivite süresi ortalamaları karşılaştırıldığında, Tennis aktivitesinde en yüksek ortalama değere (yaklaşık 0.1) ulaşıldığı, bunu Running aktivitesinin (yaklaşık 0.07) takip ettiği gözlemlenmiştir. En düşük ortalama değer ise Tennis'in son grubunda (yaklaşık -0.1) kaydedilmiştir. Güven aralıklarının önemli ölçüde örtüşmesi, aktivite türleri arasındaki farklılıkların istatistiksel olarak anlamlı olmayabileceğini göstermektedir. Bu görsel bulgu, önceki analizlerde elde edilen aktivite türleri arasında anlamlı bir fark olmadığı sonucunu desteklemektedir.



Şekil 36 Farklı Aktivite Türlerinin Günlük Adım Sayısı Üzerindeki Etkisi

Farklı aktivite türlerinin günlük adım sayısı üzerindeki etkisini gösteren ortalama grafiği incelendiğinde (Şekil 36), aktivite türleri arasında günlük adım sayısı ortalamaları karşılaştırıldığında, Running aktivitesinde en yüksek ortalama değere (yaklaşık 0.05) ulaşıldığı, bunu Tennis ve Yoga aktivitelerinin (yaklaşık 0.03) takip ettiği gözlemlenmiştir. En düşük ortalama değer ise Basketball aktivitesinde (yaklaşık -0.05) kaydedilmiştir. Tüm aktivite türleri için güven aralıklarının önemli ölçüde örtüşmesi, gruplar arasındaki farklılıkların istatistiksel olarak anlamlı olmayabileceğini göstermektedir. Bu görsel analiz, aktivite türleri arasında günlük adım sayısı bakımından belirgin bir farklılık olmadığı bulgusunu desteklemektedir.



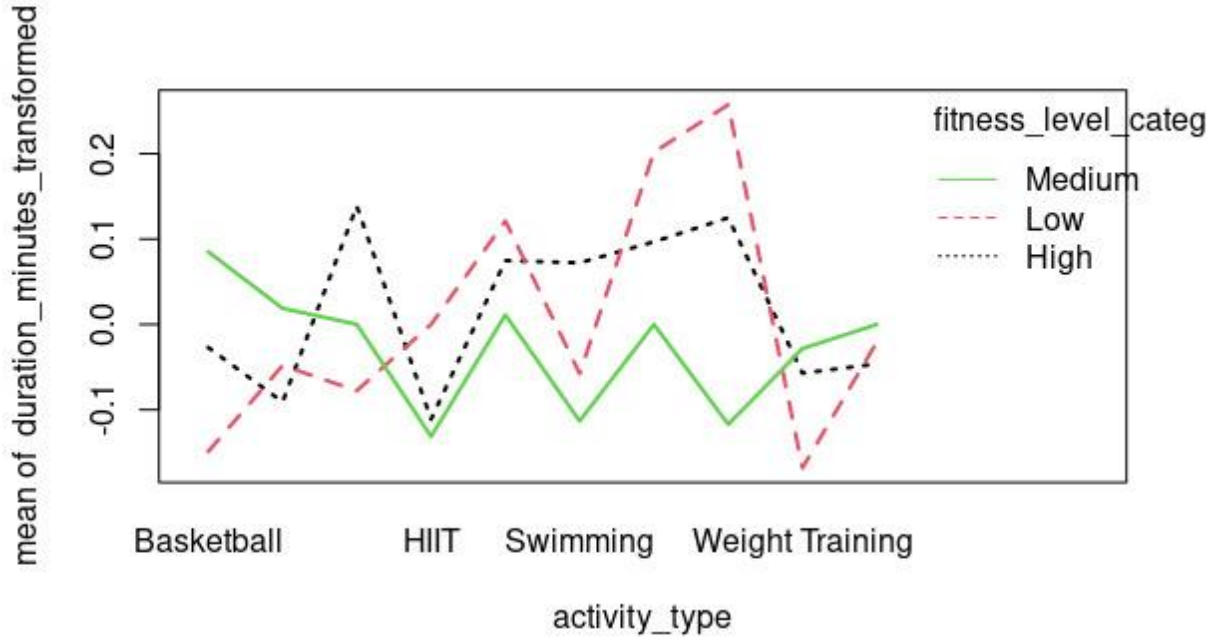
Şekil 37 Farklı Aktivite Türlerinin Günlük Adım Sayısı Üzerindeki Etkisi

Şekil 37’de farklı aktivite türlerinin günlük adım sayısı üzerindeki etkisini inceleyen grafik, aktivite türleri ile fitness seviyeleri (Low, Medium, High) arasında bir karşılaştırma sunmaktadır. Basketbol, HIIT, yüzme ve ağırlık antrenmanları gibi aktivitelerin, günlük adım sayısına etkisi fitness seviyelerine göre farklılık göstermektedir.

Medium fitness seviyesinde yer alan bireyler, genel olarak diğer seviyelere göre daha yüksek ortalama değerlere sahiptir. Low fitness seviyesinde yer alan bireyler, özellikle yüzme ve HIIT aktivitelerinde daha yüksek günlük adım ortalamalarına yaklaşırken, ağırlık antrenmanında daha düşük değerlere sahiptir. High fitness seviyesindekiler ise tüm aktivite türlerinde diğer seviyelere göre daha düşük ortalama değerlere sahiptir.

Basketbol aktivitesinde ortalama günlük adım sayısı tüm seviyelerde en düşük düzeyde görülürken, ağırlık antrenmanında özellikle Medium fitness seviyesindekiler için belirgin bir artış gözlemlenmektedir. Bununla birlikte, tüm aktivite türlerinde çizgilerin oldukça yakın seyretmesi ve güven aralıklarının kesişmesi, aktivite türleri ve fitness seviyeleri arasındaki farklılıkların istatistiksel olarak anlamlı olmayabileceğini göstermektedir. Fitness seviyesi High olan bireylerin günlük ortalama adım sayılarının, diğer fitness seviyelerine (Medium ve

Low) kıyasla belirgin şekilde daha düşük olduğu dikkat çekmektedir. Bu durum, bu grup ile diğer gruplar arasındaki farklılığın ayrıştırılabilir bir düzeyde olduğunu göstermektedir.



Farklı aktivite türlerinin ortalama süre değerleri üzerindeki etkisini gösteren grafikte (Şekil 38), Low fitness seviyesine sahip bireylerin ağırlık antrenmanında en yüksek ortalama süreye ulaştığı, buna karşılık yüzme ve HIIT aktivitelerinde diğer gruplara kıyasla daha belirgin dalgalanmalar sergilediği görülmektedir.

High fitness seviyesindeki bireyler için, süre ortalamalarının özellikle basketbol ve HIIT aktivitelerinde diğer gruplardan ayrılarak daha yüksek olduğu dikkat çekmektedir. Buna karşın, yüzme ve ağırlık antrenmanlarında Medium fitness seviyesindeki bireylerin ortalama süre değerleri daha düşük bir eğilim göstermektedir.

Tüm fitness seviyeleri için bazı aktivite türlerinde güven aralıklarının örtüşmesi, gruplar arası farklılıkların bazı durumlarda istatistiksel olarak anlamlı olmayabileceğine işaret etmektedir. Ancak, belirli aktivitelerde görülen bu farklılıklar, fitness seviyelerinin aktivite süreleri üzerindeki etkisini incelemek açısından önemli bir bulgu olarak düşünülebilir.

6. Temel Bileşenler ve Faktör Analizi

Çok değişkenli veri setlerinin analizi, yüksek boyutlu verilerdeki ilişkileri anlamak, boyut azaltma ve görselleştirme süreçlerini etkin bir şekilde yönetmek için kritik bir öneme sahiptir. Temel Bileşenler Analizi (TBA) ve Faktör Analizi (FA), bu tür veri setlerinde sıklıkla kullanılan yöntemlerdir. TBA, değişkenler arasındaki korelasyonları analiz ederek toplam varyansı açıklayan ve birbirinden bağımsız yeni bileşenler oluşturmayı hedefler. FA ise veri setindeki ortak varyansı açıklamak için gizli faktörleri belirler ve değişkenler arasındaki temel yapıyı ortaya koyar.

Bu çalışmada TBA ve FA, sağlık ve fitness verilerinin analitik süreçlerinde önemli bir rol oynayacaktır. FitnessApp veri setindeki 19 değişken arasındaki ilişkilerin incelenmesiyle, boyut azaltma işlemleri gerçekleştirilerek verinin analitik olarak daha yönetilebilir hale getirilmesi sağlanacaktır. Özellikle, fitness seviyesi, günlük adım sayısı, stres seviyesi ve hidrasyon seviyesi gibi değişkenlerin ortak varyansını açıklayan faktörlerin belirlenmesi, yaşam tarzı ve sağlık durumu arasındaki ilişkilerin daha net bir şekilde anlaşılmasına olanak tanıyacaktır.

Bu yöntemlerin uygulanmasıyla veri setinin karmaşıklığı azalacak, daha hızlı ve verimli analizler yapılabilecektir.

Projedeki örneklem büyüklüğünün 3000 olması, faktör analizinde kullanılacak faktör yükü eşik değerinin belirlenmesinde önemli bir rol oynamaktadır. Literatürde yaygın olarak kabul edilen kriterlere göre, örneklem büyüklüğüne bağlı olarak minimum faktör yükü eşikleri önerilmektedir. Örneklem büyüklüğü 3000 gibi geniş bir veri seti için faktör yükü eşik değeri genellikle 0,30 olarak kabul edilir. Bu değer, faktörlerin belirli bir değişken üzerindeki etkisini anlamlı bir şekilde yansıttığını ve güvenilir sonuçlar elde edilmesini sağladığını göstermektedir. Bu bağlamda, analizlerde faktör yükü değerleri için 0,30 eşiğinin kullanılması, veri setindeki değişkenlerin faktör yapısına uygunluğunu değerlendirmek açısından yeterli bir duyarlılığı temsil etmektedir.

Correlation Matrix ^a														
		hydration_level_transformed	calories_burned_transformed	bmi_transformed	fitness_level_transformed	age_transformed	duration_minutes_transformed	avg_heart_rate_transformed	hours_sleep_transformed	resting_heart_rate_transformed	stress_level_transformed	daily_steps_transformed	blood_pressure_systolic_transformed	blood_pressure_diastolic_transformed
Correlation	hydration_level_transformed	1.000	-.007	-.018	-.015	.005	.010	-.002	.037	.014	.005	-.006	.003	-.009
	calories_burned_transformed	-.007	1.000	.319	.291	.002	.686	.099	.013	.007	-.026	-.110	.015	-.048
	bmi_transformed	-.018	.319	1.000	.846	.031	-.021	-.036	-.014	-.009	-.014	-.356	.005	-.009
	fitness_level_transformed	-.015	.291	.846	1.000	.004	.008	-.008	-.009	.019	-.035	-.369	-.008	-.011
	age_transformed	.005	.002	.031	.004	1.000	-.004	-.566	.010	-.006	.031	-.005	-.014	-.012
	duration_minutes_transformed	.010	.686	-.021	.008	-.004	1.000	.006	.010	.018	-.018	-.015	.027	-.037
	avg_heart_rate_transformed	-.002	.099	-.036	-.008	-.566	.006	1.000	.023	.011	-.038	.001	.019	-.009
	hours_sleep_transformed	.037	.013	-.016	-.009	.010	.010	.023	1.000	-.026	-.016	-.029	.018	.000
	resting_heart_rate_transformed	.014	.007	.009	.019	-.006	.018	.011	-.026	1.000	-.015	-.005	-.009	.036
	stress_level_transformed	.005	-.026	-.014	-.035	.031	-.018	-.038	-.016	-.015	1.000	.042	-.015	-.016
	daily_steps_transformed	-.006	-.110	-.356	-.369	-.005	-.015	.001	-.029	-.005	.042	1.000	.020	.019
	blood_pressure_systolic_transformed	.003	.015	.005	-.008	-.014	.027	.019	.018	-.009	-.015	.020	1.000	.001
	blood_pressure_diastolic_transformed	-.009	-.048	-.009	-.011	-.012	-.037	-.009	.000	.036	-.016	.019	.001	1.000
Sig. (1-tailed)	hydration_level_transformed		.355	.163	.208	.385	.298	.465	.021	.223	.389	.371	.445	.313
	calories_burned_transformed	.355		.000	.000	.454	.000	.000	.245	.344	.077	.000	.203	.004
	bmi_transformed	.163	.000		.000	.045	.130	.026	.198	.315	.214	.000	.395	.305
	fitness_level_transformed	.208	.000	.000		.406	.322	.340	.303	.146	.027	.000	.330	.279
	age_transformed	.385	.454	.045	.406		.409	.000	.297	.372	.042	.395	.220	.261
	duration_minutes_transformed	.298	.000	.130	.322	.409		.368	.293	.159	.160	.201	.073	.022
	avg_heart_rate_transformed	.465	.000	.026	.340	.000	.368		.099	.273	.018	.485	.149	.318
	hours_sleep_transformed	.021	.245	.198	.303	.297	.293	.099		.078	.184	.056	.162	.500
	resting_heart_rate_transformed	.223	.344	.315	.146	.372	.159	.273	.078		.205	.401	.320	.025
	stress_level_transformed	.389	.077	.214	.027	.042	.160	.018	.184	.205		.010	.206	.186
	daily_steps_transformed	.371	.000	.000	.000	.395	.201	.485	.056	.401	.010		.138	.151
	blood_pressure_systolic_transformed	.445	.203	.395	.330	.220	.073	.149	.162	.320	.206	.138		.483
	blood_pressure_diastolic_transformed	.313	.004	.305	.279	.261	.022	.318	.500	.025	.186	.151	.483	

a. Determinant = .065

Şekil 38 Korelasyon Matrisi

Şekil 38’de aktarılan korelasyon matrisi, farklı sağlık ve fitness değişkenleri arasındaki ilişkileri göstermektedir. Egzersiz süresi ve yakılan kalori miktarı arasında yüksek bir pozitif korelasyon ($r=0.686$) vardır ve bu ilişki istatistiksel olarak anlamlıdır ($p < 0.001$). Bu, daha uzun süre egzersiz yapmanın daha fazla kalori yakılmasıyla ilişkilendirildiğini göstermektedir.

Ortalama kalp atış hızı ve yaş arasında negatif bir korelasyon ($r=-0.566$) bulunmakta, yaş arttıkça ortalama kalp hızının azaldığını işaret etmektedir. Bu ilişkinin de oldukça anlamlı olduğu söylenebilir ($p < 0.001$).

Inverse of Correlation Matrix														
		hydration_level_transformed	calories_burned_transformed	bmi_transformed	fitness_level_transformed	age_transformed	duration_minutes_transformed	avg_heart_rate_transformed	hours_sleep_transformed	resting_heart_rate_transformed	stress_level_transformed	daily_steps_transformed	blood_pressure_systolic_transformed	blood_pressure_diastolic_transformed
hydration_level_transformed		1.002	.019	.013	.003	-.007	-.021	-.003	-.037	-.015	-.006	.012	-.002	.009
calories_burned_transformed		.019	2.495	-.844	.021	-.200	-1.728	-.379	-.019	.022	.016	-.062	.018	.044
bmi_transformed		.013	-.844	3.858	-2.948	.012	.684	.198	.037	.015	-.059	.206	-.057	-.015
fitness_level_transformed		.003	-.021	-2.948	3.605	.050	-.071	-.045	-.003	-.039	.068	.275	.042	.005
age_transformed		-.007	-.200	.012	.050	1.491	.139	.864	-.033	-.004	-.015	.008	.004	.021
duration_minutes_transformed		-.021	-1.728	.684	-.071	.139	2.203	.261	.005	-.034	.005	.061	-.041	.007
avg_heart_rate_transformed		-.003	-.379	.198	-.045	.864	.261	1.534	-.039	-.016	.026	.017	-.019	.017
hours_sleep_transformed		-.037	-.019	.037	-.003	-.033	.005	-.039	1.006	.027	.015	.039	-.018	-.003
resting_heart_rate_transformed		-.015	.022	.015	-.039	-.004	-.034	-.016	.027	1.004	.014	-.002	.009	-.036
stress_level_transformed		-.006	.016	-.059	.068	-.015	.005	.026	.015	.014	1.006	-.037	.015	.018
daily_steps_transformed		.012	-.062	.206	.275	.008	.061	.017	.039	-.002	-.037	1.172	-.024	-.018
blood_pressure_systolic_transformed		-.002	.018	-.057	.042	.004	-.041	-.019	-.018	.009	.015	-.024	1.003	-.001
blood_pressure_diastolic_transformed		.009	.044	-.015	.005	.021	.007	.017	-.003	-.036	.018	-.018	-.001	1.005

Şekil 39 Korelasyon Matrisinin Tersi

VIF (Variance Inflation Factor) değerinin 10'un üzerinde olması, genellikle çoklu bağlantı probleminin varlığına işaret eder. Ancak, verilen korelasyon matrisinin tersine bakıldığında, köşegen üzerindeki değerlerin hiçbirinin 5'in üzerinde olmadığı gözlemlenmiştir. Bu durum, değişkenler arasında ciddi bir çoklu bağlantı probleminin olmadığını göstermektedir. Dolayısıyla, değişkenler arasındaki bağımlılık seviyesinin kabul edilebilir sınırlar içinde olduğu ve regresyon analizlerinin güvenilir bir şekilde gerçekleştirilebileceği sonucuna varılabilir.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.533
Bartlett's Test of Sphericity	Approx. Chi-Square	8196.212
	df	78
	Sig.	<.001

Şekil 40 KMO ve Bartlett Testleri

Faktör analizi öncesi verilerin uygunluğunun değerlendirilmesi amacıyla gerçekleştirilen (KMO) örneklem yeterliliği testi ve Bartlett küresellik testi sonuçları incelenmiştir. KMO değeri 0.533 olarak hesaplanmış olup, bu değer örneklem büyüklüğünün faktör analizi için zayıf düzeyde yeterli olduğunu göstermektedir.

H_0 : Korelasyon matrisi birim matristir (değişkenler arasında ilişki yoktur).

H_1 : Korelasyon matrisi birim matris değildir (değişkenler arasında ilişki vardır).

Bartlett küresellik testi sonuçları ise H_0 reddedildiğinden istatistiksel olarak anlamlı bulunmuştur ($\chi^2=8196.212$, $df=78$, $p<0.001$). Bartlett testi sonucunun anlamlı çıkması, değişkenler arasında faktör analizi için yeterli düzeyde ilişki olduğunu gösterir.

Anti-image Matrices														
		hydration_level_transformed	calories_burned_transformed	bmi_transformed	fitness_level_transformed	age_transformed	duration_minutes_transformed	avg_heart_rate_transformed	hours_sleep_transformed	resting_heart_rate_transformed	stress_level_transformed	daily_steps_transformed	blood_pressure_systolic_transformed	blood_pressure_diastolic_transformed
Anti-image Covariance	hydration_level_transformed	.998	.008	.003	.001	-.005	-.010	-.002	-.037	-.015	-.006	.011	-.002	.009
	calories_burned_transformed	.008	.401	-.088	-.002	-.054	-.314	-.099	-.007	.009	.006	-.021	.007	.018
	bmi_transformed	.003	-.088	.259	-.212	.002	.081	.034	.010	.004	-.015	.046	-.015	-.004
	fitness_level_transformed	.001	-.002	-.212	.277	.009	-.009	-.008	-.001	-.011	.019	.065	.012	.001
	age_transformed	-.005	-.054	.002	.009	.671	.042	.378	-.022	-.003	-.010	.005	.003	.014
	duration_minutes_transformed	-.010	-.314	.081	-.009	.042	.454	.077	.002	-.015	.002	.024	-.018	.003
	avg_heart_rate_transformed	-.002	-.099	.034	-.008	.378	.077	.652	-.025	-.010	.017	.010	-.012	.011
	hours_sleep_transformed	-.037	-.007	.010	-.001	-.022	.002	-.025	.994	.027	.015	.033	-.018	-.003
	resting_heart_rate_transformed	-.015	.009	.004	-.011	-.003	-.015	-.010	.027	.996	.013	-.002	.009	-.036
	stress_level_transformed	-.006	.006	-.015	.019	-.010	.002	.017	.015	.013	.994	-.031	.015	.017
	daily_steps_transformed	.011	-.021	.046	.065	.005	.024	.010	.033	-.002	-.031	.853	-.021	-.015
	blood_pressure_systolic_transformed	-.002	.007	-.015	.012	.003	-.018	-.012	-.018	.009	.015	-.021	.997	-.001
	blood_pressure_diastolic_transformed	.009	.018	-.004	.001	.014	.003	.011	-.003	-.036	.017	-.015	-.001	.995
	Anti-image Correlation	hydration_level_transformed	.517*	.012	.007	.001	-.006	-.014	-.002	-.037	-.015	-.006	.012	-.002
calories_burned_transformed		.012	.505*	-.272	-.007	-.104	-.737	-.194	-.012	.014	.010	-.036	.011	.028
bmi_transformed		.007	-.272	.551*	-.791	.005	.235	.082	.019	.008	-.030	.097	-.029	-.007
fitness_level_transformed		.001	-.007	-.791	.592*	.021	-.025	-.019	-.002	-.020	.035	.134	.022	.003
age_transformed		-.006	-.104	.005	.021	.484*	.077	.571	-.027	-.004	-.012	.006	.004	.017
duration_minutes_transformed		-.014	-.737	.235	-.025	.077	.430*	.142	.003	-.023	.003	.038	-.027	.005
avg_heart_rate_transformed		-.002	-.194	.082	-.019	.571	.142	.459*	-.031	-.013	.021	.013	-.015	.013
hours_sleep_transformed		-.037	-.012	.019	-.002	-.027	.003	-.031	.436*	.027	.015	.036	-.018	-.003
resting_heart_rate_transformed		-.015	.014	.008	-.020	-.004	-.023	-.013	.027	.470*	.014	-.002	.009	-.036
stress_level_transformed		-.006	.010	-.030	.035	-.012	.003	.021	.015	.014	.606*	-.034	.015	.018
daily_steps_transformed		.012	-.036	.097	.134	.006	.038	.013	.036	-.002	-.034	.892*	-.022	-.017
blood_pressure_systolic_transformed		-.002	.011	-.029	.022	.004	-.027	-.015	-.018	.009	.015	-.022	.420*	-.001
blood_pressure_diastolic_transformed		.009	.028	-.007	.003	.017	.005	.013	-.003	-.036	.018	-.017	-.001	.646*

a. Measures of Sampling Adequacy(MSA)

Şekil 41 Anti-Image-Korelasyon Matrisi

Anti-image-korelasyon matrisini incelediğinde (Şekil 41), değişkenler arasındaki ilişkilerin yapısını gözlemleyebiliriz. Matriste köşegen elemanları MSA (Measure of Sampling Adequacy - Örneklem Yeterliliği Ölçüsü) değerlerini göstermektedir. Analizde dikkat çeken önemli bir nokta; yaş, aktivite süresi, ortalama kalp atış hızı, uyku süresi, dinlenme anı kalp atış hızı, kan basıncı (sistolik) değişkenlerinin MSA değerinin 0.50'nin altında olmasıdır. Bu durum normalde söz konusu değişkenlerin analizden çıkarılmasını gerektirebilir, ancak çalışmadaki değişken sayısının yeterliliği göz önünde bulundurularak faktör analizine devam edilmesi uygun görülmüştür. Diğer değişkenlerin MSA değerleri kabul edilebilir sınırlar içerisinde. Matristeki diğer elemanlar, değişkenler arasındaki anti-image-korelasyon katsayılarını göstermekte olup, bu değerler faktör analizinin uygunluğunu değerlendirmede önemli rol oynamaktadır.

Communalities

	Initial	Extraction
hydration_level_transformed	1.000	.680
calories_burned_transformed	1.000	.867
bmi_transformed	1.000	.857
fitness_level_transformed	1.000	.855
age_transformed	1.000	.782
duration_minutes_transformed	1.000	.874
avg_heart_rate_transformed	1.000	.786
hours_sleep_transformed	1.000	.550
resting_heart_rate_transformed	1.000	.595
stress_level_transformed	1.000	.408
daily_steps_transformed	1.000	.390
blood_pressure_systolic_transformed	1.000	.441
blood_pressure_diastolic_transformed	1.000	.440

Extraction Method: Principal Component Analysis.

Şekil 42 Ortaklık Grafiği

Şekil 42, yapılan temel bileşen analizi sonucunda her bir değişken için başlangıçtaki (Initial) ve çıkarılan (Extraction) ortaklık değerlerini göstermektedir. Analiz sonucunda, değişkenlerin en az %50'sinin (0.5) açıklanması gerekmektedir. Bu kritere göre değerlendirildiğinde, bazı değişkenlerin açıklanan ortaklık değerlerinin (%50'den düşük olanlar) yeterli olmadığı görülmüştür.

Özellikle, stress_level_transformed (0.408), daily_steps_transformed (0.390), blood_pressure_systolic_transformed (0.441) ve blood_pressure_diastolic_transformed (0.440) değişkenlerinin değerleri 0.5'in altındadır. Bu durum, bu değişkenlerin analize anlamlı bir katkı sağlamadığını ve çıkarılmalarının uygun olabileceğini göstermektedir.

Düşük ortaklık değerine sahip değişkenler analizden çıkarıldıktan sonra elde edilen yeni tablo, Şekil 43'te sunulmuştur.

Communalities

	Initial	Extraction
hydration_level_transformed	1.000	.673
calories_burned_transformed	1.000	.872
bmi_transformed	1.000	.925
fitness_level_transformed	1.000	.912
age_transformed	1.000	.782
duration_minutes_transformed	1.000	.889
avg_heart_rate_transformed	1.000	.787
hours_sleep_transformed	1.000	.604
resting_heart_rate_transformed	1.000	.777

Extraction Method: Principal Component Analysis.

Şekil 43 Ortaklık Grafiği-2

Şekil 43, düşük ortaklık değerine sahip değişkenlerin analizden çıkarılması sonrası elde edilen güncel durumu göstermektedir. Bu işlem sonucunda, tüm değişkenlerin ortaklık değerlerinin (%50'nin üzerinde) yeterince yüksek olduğu görülmektedir.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.112	23.468	23.468	2.112	23.468	23.468	1.916	21.291	21.291
2	1.601	17.784	41.252	1.601	17.784	41.252	1.682	18.683	39.975
3	1.456	16.183	57.435	1.456	16.183	57.435	1.569	17.433	57.408
4	1.038	11.536	68.971	1.038	11.536	68.971	1.037	11.520	68.928
5	1.013	11.254	80.224	1.013	11.254	80.224	1.017	11.296	80.224
6	.947	10.524	90.748						
7	.444	4.939	95.687						
8	.244	2.708	98.395						
9	.144	1.605	100.000						

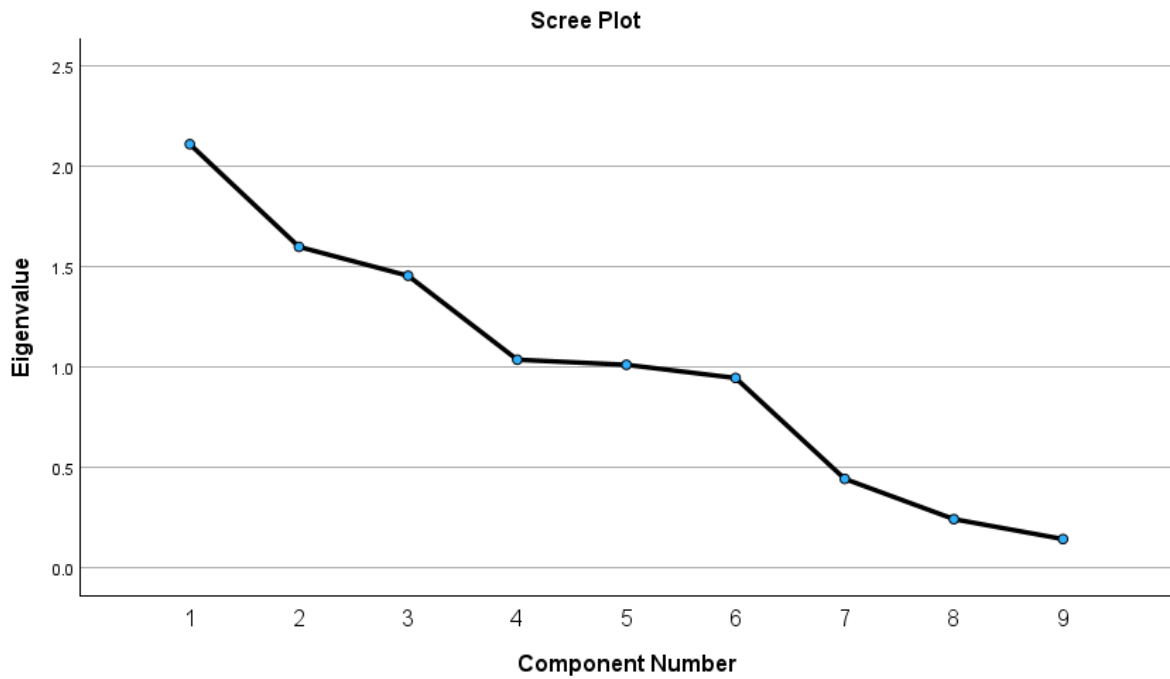
Extraction Method: Principal Component Analysis.

Şekil 44 Total Varyans Açıklama Oranları

Şekil 44'te sunulan "Toplam Açıklanan Varyans" verilerine göre, yapılan Temel Bileşenler Analizi (Principal Component Analysis - PCA) sonucunda özdeğeri (eigenvalue) 1'den büyük olan toplam beş faktör elde edilmiştir. İlk faktör toplam varyansın %21,29'unu, ikinci faktör %18,68'ini, üçüncü faktör %17,43'ünü, dördüncü faktör %11,52'sini ve beşinci faktör

%11,29'unu açıklamaktadır. Bu beş faktör bir arada toplam varyansın %80,22'sini açıklamaktadır.

Bu sonuçlar, faktörlerin veri setindeki değişkenlerin büyük bir kısmını temsil edebildiğini ve faktör sayısının yeterli olduğunu göstermektedir. Özellikle ilk iki faktör toplam varyansın yaklaşık %40'ını açıklayarak verilerin büyük bir bölümünü özetlemektedir. Dolayısıyla, faktör sayısının artırılmasına gerek olmadığı ve bu beş faktörle analizlerin güvenilir bir şekilde sürdürülebileceği sonucuna varılabilir.



Şekil 45 Scree Plot

Scree Plot grafiği (Şekil 45), faktörlerin özdeğerlerine göre sıralanarak sunulduğu bir görselleştirmedir ve faktör sayısının belirlenmesinde önemli bir araçtır. Grafikte, birinci ve ikinci bileşenlerin özdeğerlerinin 1'in üzerinde olduğu ve belirgin bir şekilde diğer bileşenlere kıyasla daha yüksek değerler aldığı gözlemlenmektedir. Beşinci bileşenden itibaren eğrideki düşüşün giderek yataylaştığı ve özdeğerlerin 1'in altına düştüğü görülmektedir. Bu durum, ilk beş faktörün veri setindeki varyansın büyük bir kısmını açıkladığını ve sonraki bileşenlerin açıklayıcı gücünün oldukça azaldığını göstermektedir. Dolayısıyla, bu grafikten hareketle beş

faktörlü bir yapı benimsemenin uygun olduğu söylenebilir. Bu faktörler, verinin temel yapısını anlamak için yeterli bir temsil sağlamaktadır.

Component Matrix^a

	Component				
	1	2	3	4	5
bmi_transformed	.832		-.405		
fitness_level_transformed	.828		-.416		
calories_burned_transformed	.732	.306	.492		
avg_heart_rate_transformed		.805	-.371		
age_transformed		-.775	.426		
duration_minutes_transformed	.443	.371	.744		
hours_sleep_transformed				.765	
hydration_level_transformed				.603	.553
resting_heart_rate_transformed					.831

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Şekil 46 Bileşen Matrisi

Şekil 46'daki bileşen matrisi tablosunda, her bir değişkenin faktörlerle olan ilişki düzeyini gösteren korelasyon katsayıları yer almaktadır. İlk faktörde (1. Bileşen) dört değişken (bmi_transformed, fitness_level_transformed, calories_burned_transformed, avg_heart_rate_transformed) yüksek yük değerleriyle temsil edilmektedir. İkinci faktörde (2. Bileşen) ise iki değişken (age_transformed, duration_minutes_transformed) öne çıkmaktadır. Korelasyon katsayılarının 0,30'un üzerinde olması, bu değişkenlerin faktörlerle anlamlı bir ilişki sergilediğini göstermektedir.

Bu katsayıların kareleri, değişkenlerin ilgili faktör tarafından açıklanan varyans yüzdelерini vermektedir. Örneğin, bmi_transformed değişkeni için ($0,832^2 = 0,692$), yani yaklaşık %69'u birinci faktör tarafından açıklanmaktadır. Ayrıca, age_transformed değişkeni ve diğer değişkenler arasındaki farkların mutlak değerce 0,1'den büyük olması, değişkenlerin faktörlerden çıkarılmasına gerek olmadığını göstermektedir.

Sonuç olarak, değişkenlerin faktör yapısı içindeki dağılımı tutarlı ve yorumlanabilir bir yapı sergilemektedir. Bir sonraki adımda Varimax döndürme yöntemi uygulanarak faktör yüklerinin daha net bir şekilde ayrıştırılması ve yorumlanabilirliğin artırılması sağlanacaktır.

Rotated Component Matrix^a

	Component				
	1	2	3	4	5
bmi_transformed	.959				
fitness_level_transformed	.952				
duration_minutes_transformed		.937			
calories_burned_transformed		.889			
avg_heart_rate_transformed			.885		
age_transformed			-.884		
hydration_level_transformed				.765	
hours_sleep_transformed				.671	-.392
resting_heart_rate_transformed					.881

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Şekil 47 Döndürülmüş Bileşen Matrisi

Döndürülmemiş bileşen matrisinde (Component Matrix), birçok değişkenin birden fazla bileşene önemli yükler verdiği gözlenmektedir. Örneğin, "bmi_transformed" ve "fitness_level_transformed" değişkenleri hem birinci bileşene hem de diğer bileşenlere yüklenmektedir. Bu durum, bileşenlerin temsil ettikleri faktörlerin yorumlanmasını zorlaştırmaktadır. Ayrıca, değişkenler arasında yüksek yük değerlerinin dağılımı düzensizdir ve bu da anlamlı faktör kümelerinin ortaya çıkmasını güçleştirir.

Döndürülmüş bileşen matrisinde (Rotated Component Matrix) ise, her bir değişkenin bir bileşene daha yüksek bir yük verirken diğer bileşenlerle olan yüklerinin minimal düzeye indiği görülmektedir. Örneğin, "bmi_transformed" ve "fitness_level_transformed" değişkenleri, açıkça birinci bileşene yüklenirken, "duration_minutes_transformed" ve "calories_burned_transformed" değişkenleri ikinci bileşene daha belirgin şekilde yüklenmiştir. Bu durum, bileşenlerin daha kolay tanımlanabilir hale gelmesini sağlar ve her bir bileşenin temsil ettiği yapının daha açık bir şekilde anlaşılmasına olanak tanır.

Döndürülmüş bileşen matrisine (Rotated Component Matrix) dayanarak, her bir bileşen için faktör yükleri kullanılarak doğrusal denklemler oluşturulabilir. Bu denklemler, her bir bileşeni açıklayan değişkenlerin katkısını (faktör yüklerini) temsil eder ve bileşenlerin matematiksel olarak ifade edilmesini sağlar.

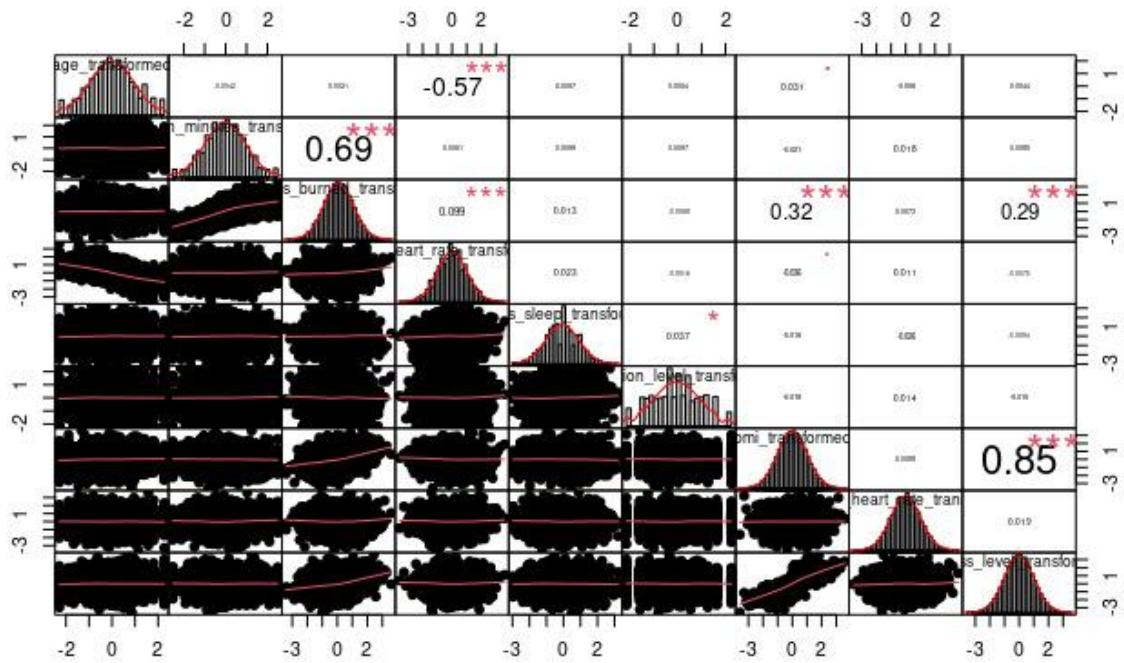
$$Y_1 = 0.959 * (\text{bmi transformed}) + 0.952 * (\text{fitness level transformed})$$

$$Y_2 = 0.937 * (\text{duration minutes transformed}) + 0.889 * (\text{calories burned transformed})$$

$$Y_3 = 0.885 * (\text{avg. heart rate transformed}) - 0.884 * (\text{age transformed})$$

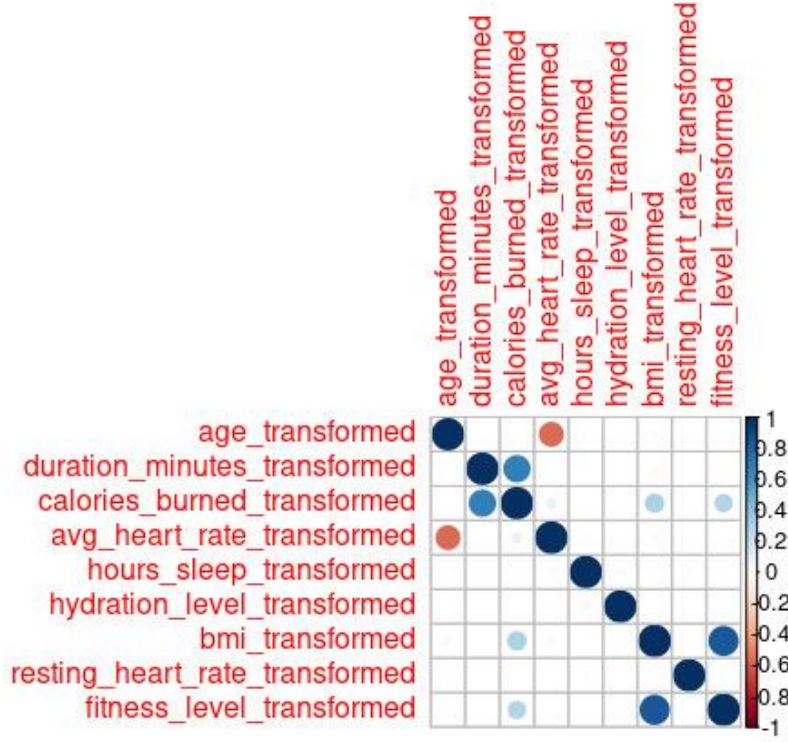
$$Y_4 = 0.765 * (\text{hydration level transformed}) + 0.671 * (\text{hours sleep transformed})$$

$$Y_5 = -0.392 * (\text{hours sleep transformed}) + 0.881 * (\text{resting heart rate transformed})$$



Şekil 48 Korelasyon Matrisi ve Dağılımlar

Şekil 48’de, yıldız sembolleri ile belirtilen katsayılar, istatistiksel olarak anlamlı ilişkilere işaret etmektedir. Genellikle, **, *** gibi yıldızlar, istatistiksel anlamlılık seviyesini gösterir ve bu seviyeler, $p < 0.01$ veya $p < 0.001$ gibi belirli bir güven seviyesinde bu ilişkinin şans eseri olmadığını belirtir. Korelasyon matrisinde yıldızla gösterilen değişkenler anlamlı ilişkiye sahip değişkenlerdir ve bu değişkenlerin belirli bileşenler altında toplanır.



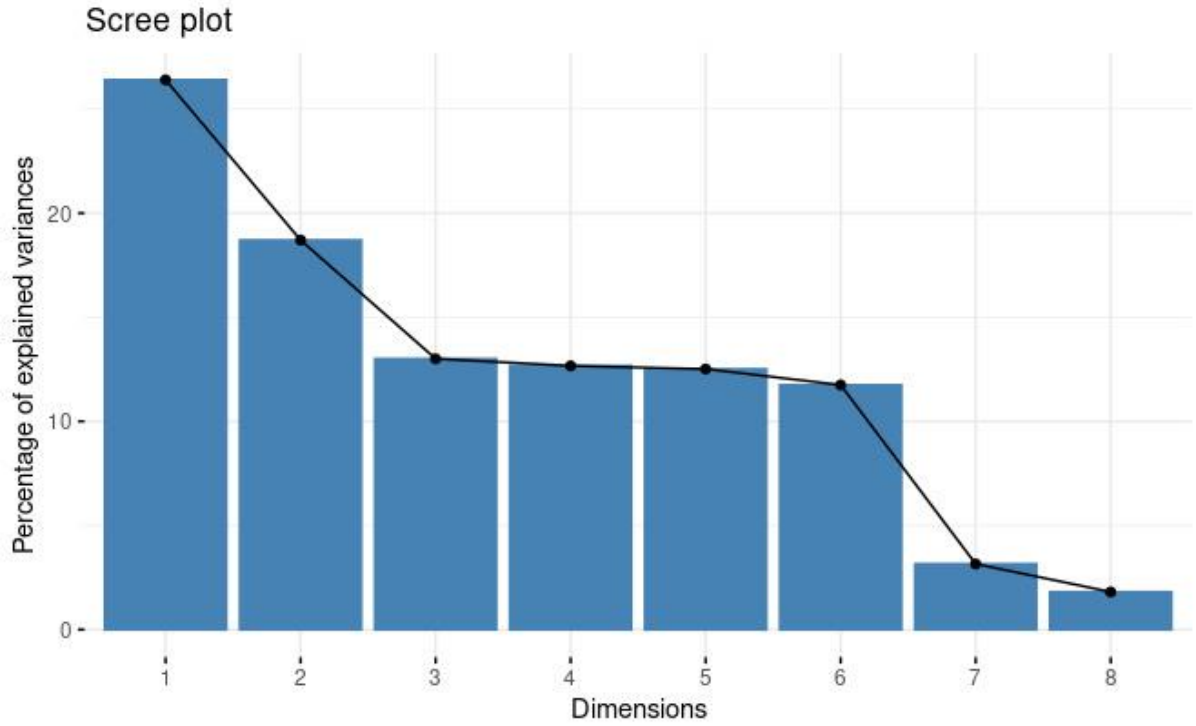
Şekil 49 Korelasyon Matrisi ve Daireler

Renk kodlaması, korelasyonun yönü ve büyüklüğü hakkında hızlı bir görsel referans sağlar (Şekil 49). Mavi tonlar negatif korelasyonu, turuncu tonlar ise pozitif korelasyonu temsil eder. Renklerin yoğunluğu, korelasyonun gücünü yansıtır; daha koyu mavi renkler daha düşük, daha koyu turuncu renkler ise daha yüksek korelasyonu gösterir.

Örneğin:

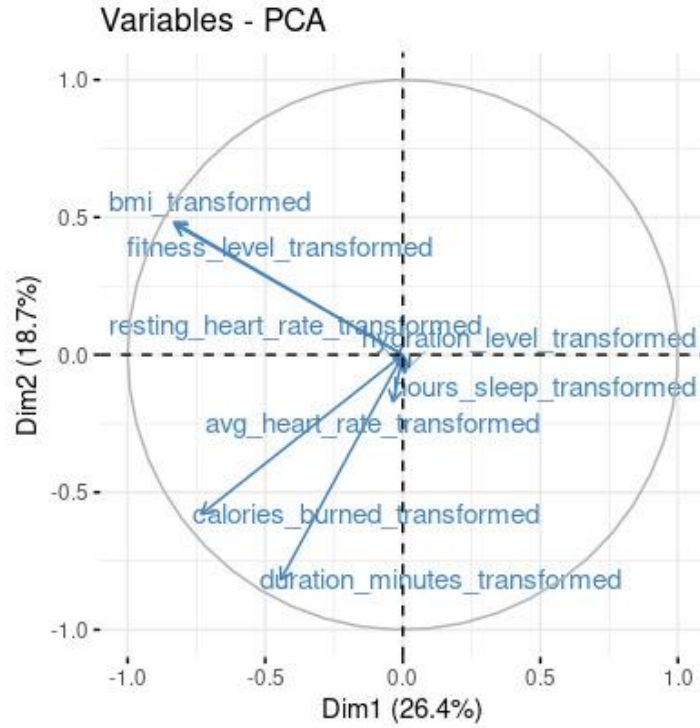
- **Mavi tonlar**, negatif bir ilişkiyi temsil eder. Bir değişken artarken diğerinin azaldığını ifade eder. Renk ne kadar koyu mavi olursa, bu negatif ilişki o kadar güçlüdür.
- **Turuncu tonlar**, pozitif bir ilişkiyi gösterir. Bir değişkenin artışıyla diğerinin de artma eğiliminde olduğunu ifade eder. Renk ne kadar koyu turuncu olursa, bu pozitif ilişki o kadar güçlüdür.

Bu görseldeki korelasyon matrisi, örneğin "avg_heart_rate_transformed" ile "calories_burned_transformed" arasında koyu turuncu bir kare gösteriyorsa, bu değişkenler arasında güçlü pozitif bir ilişki olduğunu belirtir. Benzer şekilde, "age_transformed" ile "hydration_level_transformed" arasında açık mavi bir renk varsa, bu değişkenler arasında zayıf negatif bir ilişki olduğuna işaret eder.



Şekil 50 Scree Plot

Şekil 50'deki ilk birkaç bileşen, toplam varyansın büyük bir kısmını açıklamakta, bu da bu bileşenlerin veri setindeki değişkenlik açısından önemli olduğunu göstermektedir. Özellikle birinci bileşen, yaklaşık %25'lik bir varyans açıklamasıyla en yüksek katkıyı sunmaktadır. İkinci bileşen ise yaklaşık %20 oranında bir varyansı açıklamaktadır. Üçüncü bileşenden itibaren bileşenlerin açıklanan varyansa katkıları belirgin şekilde azalmaya başlıyor ve altıncı bileşenden sonra katkı oldukça küçük bir düzeye düşmektedir. Bu durum, boyut indirgeme işlemi için ilk beş bileşenin seçilmesinin yeterli olabileceğine işaret ediyor.



Şekil 51 İki Boyutlu Değişken Çemberi

Şekil 51'deki grafik, temel bileşen analizi (PCA) sonuçlarına göre değişkenlerin iki boyutlu düzlemde vektörlerle ifade edildiği bir değişken çemberini göstermektedir. Bu grafikte, Dim1 (Birinci Temel Bileşen) toplam varyansın %26.4'ünü, Dim2 (İkinci Temel Bileşen) ise %18.7'sini açıklamaktadır. Birinci bileşen, değişkenler arasındaki en yüksek ortak varyansı temsil ederken, ikinci bileşen birinci bileşene dik (ortogonal) olacak şekilde oluşturulmuş ve birinci bileşen tarafından açıklanamayan ek varyansı açıklamaktadır.

Değişkenler, vektörlerin yönleri ve uzunlukları üzerinden değerlendirilmiştir. Örneğin, "calories_burned_transformed" ve "duration_minutes_transformed" değişkenlerinin vektörleri benzer yönelim göstermekte ve Dim1 üzerinde güçlü bir katkıya sahip olduklarını ortaya koymaktadır. Benzer şekilde, "bmi_transformed" ve "fitness_level_transformed" değişkenleri birbirine yakın konumlanmış olup bu değişkenler arasında bir ilişki olduğunu ve birlikte hareket ettiklerini göstermektedir. "resting_heart_rate_transformed", "avg_heart_rate_transformed" ve "hours_sleep_transformed" değişkenleri ise başka bir gruba yakın konumlanmış olup bu değişkenlerin Dim2 üzerinde daha anlamlı bir katkı sağladıkları gözlemlenmektedir.

Bu analiz, deęiřkenler arasındaki iliřkileri zetleyerek, en fazla varyansı aıklayan bileřenlerin oluřturulmasını saęlamıřtır. Birinci bileřenin deęiřkenler arasındaki temel iliřkileri temsil ettięi, ikinci bileřenin ise birinci bileřen tarafından aıklanamayan varyansı yakaladıęı grlmektedir. Bu sonular, PCA'nın veri boyutunu indirgeme ve karmařık veri setlerini daha anlařılır hale getirme yeteneęini vurgulamaktadır.

7. Diskriminant Analizi

Diskriminant analizi, çok deęişkenli istatistiksel yöntemler arasında yer alan ve gruplar arası farklılıkları incelemek için kullanılan güçlü bir sınıflandırma tekniğidir. Bu yöntem, belirli deęişkenlere dayanarak gözlemleri önceden tanımlanmış gruplara atamayı veya bu gruplar arasındaki ayrışmayı deęerlendirmeyi amaçlar. Özellikle, bağımlı deęişkenin kategorik ve bağımsız deęişkenlerin sürekli olduğu durumlarda sıklıkla tercih edilir.

Diskriminant analizinin temel amacı, gruplar arasındaki en iyi ayrımı sağlayan bir doğrusal kombinasyonu belirlemektir. Bu kombinasyon, gruplar arasındaki varyansı maksimize ederken, grup içi varyansı minimize etmeye çalışır. Bu bağlamda, analiz sonucunda elde edilen diskriminant fonksiyonları, hem gruplar arası ayrımın derecesini ortaya koyar hem de gözlemlerin hangi gruba ait olduğunu tahmin etmek için bir model sağlar.

Bu çalışmada, diskriminant analizi, sağlık ve fitness verilerinden oluşan geniş bir veri seti üzerinde uygulanmıştır. Analizin odak noktası, farklı fitness seviyesi kategorileri arasındaki demografik ve yaşam tarzı faktörlerinin ayrım gücünü incelemek ve bu farklılıkların belirlenmesinde etkili olan deęişkenleri ortaya koymaktır. Bu noktada, bireylerin sağlık seviyelerini daha iyi anlamak hedeflenmiştir.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
age_transformed	1.000	.012	2	2997	.988
duration_minutes_transformed	.999	2.201	2	2997	.111
calories_burned_transformed	.926	119.505	2	2997	<.001
avg_heart_rate_transformed	1.000	.651	2	2997	.522
hours_sleep_transformed	1.000	.082	2	2997	.921
stress_level_transformed	.999	1.959	2	2997	.141
daily_steps_transformed	.757	480.795	2	2997	<.001
hydration_level_transformed	.999	2.206	2	2997	.110
bmi_transformed	.358	2684.885	2	2997	<.001
resting_heart_rate_transformed	.999	1.386	2	2997	.250
blood_pressure_systolic_transformed	1.000	.481	2	2997	.618
blood_pressure_diastolic_transformed	.999	1.799	2	2997	.166

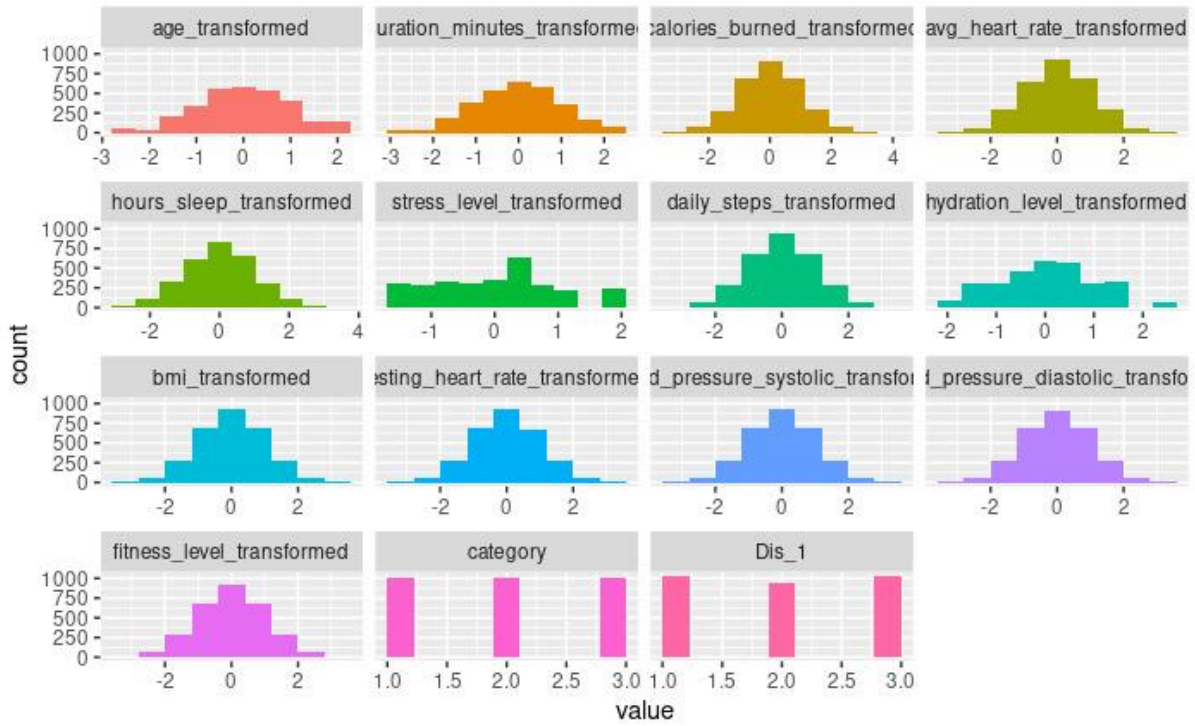
Şekil 52 Grup Ortalamaları Eşitliği Testi

Şekil 52'deki Wilks' Lambda istatistiği, her bir bağımlı değişken için grup farklılıklarının genel model üzerindeki katkısını yansıtmaktadır. Wilks' Lambda değeri ne kadar küçükse, ilgili bağımlı değişken model için o kadar fazla katkı sağlamaktadır.

Tabloda, "calories_burned_transformed", "daily_steps_transformed" ve "bmi_transformed" değişkenlerinin Wilks' Lambda değerlerinin düşük ve F değerlerinin yüksek olması, bu değişkenlerin gruplar arasında anlamlı farklılıklar gösterdiğini ortaya koymaktadır ($p < 0.001$). Bu, bu değişkenlerin modelin açıklayıcı gücüne önemli bir katkı sağladığını göstermektedir.

Diğer değişkenler ise Wilks' Lambda değerlerinin yüksek olması ve p değerlerinin 0.05'ten büyük olması nedeniyle gruplar arasında anlamlı farklılıklar göstermemektedir. Özellikle, "age_transformed", "avg_heart_rate_transformed", "hours_sleep_transformed", "stress_level_transformed", "hydration_level_transformed", "resting_heart_rate_transformed", "blood_pressure_systolic_transformed" ve "blood_pressure_diastolic_transformed" değişkenleri, modelin genel varyans açıklama gücüne daha az katkı sağlamaktadır.

Bu sonuçlar, "calories_burned_transformed", "daily_steps_transformed" ve "bmi_transformed" değişkenlerinin gruplar arası farkların belirlenmesinde daha etkili olduğunu, diğer değişkenlerin ise bu farkları açıklamada sınırlı bir rol oynadığını göstermektedir.



Şekil 53 Histogram Grafikleri

Verilen grafikler (Şekil53), nicel değişkenlerin histogramlarını göstermektedir. "fitness_level_transformed" değişkeni normal dağılıma daha yakın bir görünüm sunmaktadır.

Kategorik değişken olan "category" ve "Dis_1" değişkenlerinin histogramlarında, değerlerin belirli kategoriler etrafında yoğunlaştığı gözlenmektedir. Ancak, kategorik değişkenlerin histogramları, frekansların kategoriye göre dağılımını göstermekle sınırlıdır ve bu tür değişkenler için sürekli bir dağılımı değerlendirmek yerine, kategorilerin frekanslarını yorumlamak daha anlamlıdır. Dolayısıyla, bu tür histogramları sürekli değişkenlerin dağılımını incelemek amacıyla yorumlamak uygun olmayacaktır.

Test Results		
Box's M		146.780
F	Approx.	12.214
	df1	12
	df2	43527943.319
	Sig.	<.001
Tests null hypothesis of equal population covariance matrices.		

Şekil 54 Box' M Test Sonuçları

Şekil 54'te Box's M testi sonuçları gösterilmektedir. Test, gruplar arasında varyans-kovaryans matrislerinin eşit olup olmadığını değerlendirmek için kullanılır.

H_0 : Gruplar arası varyans – kovaryans matrisleri eşittir.
H_1 : Gruplar arası varyans – kovaryans matrisleri eşit değildir.

Tabloda verilen anlamlılık değeri (Sig.) 0.001, belirlenen anlamlılık düzeyi olan $\alpha = 0.05$ 'ten küçüktür. Bu sonuca göre, H_0 hipotezini reddedilir ve gruplar arası varyans-kovaryans matrislerinin eşit olmadığı sonucuna varılır. Bu da varsayımın sağlanamadığını göstermektedir. Box's M testinin sonucuna göre, varyans-kovaryans matrislerinin eşitliği varsayımı bu veri seti için geçerli değildir.

Normalde değişkenlerin çıkarılması gerekse de, değişkenler çıkarıldığında anlamlılık değerinin 0.05'e yaklaşmadığı gözlemlendiğinden, analiz değişkenler çıkarılmadan devam ettirilmiştir.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.103 ^a	98.1	98.1	.823
2	.041 ^a	1.9	100.0	.199

a. First 2 canonical discriminant functions were used in the analysis.

Şekil 55 Öz Değerler Tablosu

Şekil 55'teki ilk fonksiyonun özdeğeri 2.103 olup, bu değer 0.4 eşik değerinin üzerindedir. Bu fonksiyon, varyansın %98.1'ini açıklamaktadır. Kanonik korelasyon değerinin karesi (0.823^2) yaklaşık 0.677'dir.

İkinci fonksiyonun özdeğeri 0.041 olup, eşik değerin altındadır. Bu fonksiyon varyansın sadece %1.9'unu açıklamaktadır. Kanonik korelasyon değerinin karesi (0.199^2) yaklaşık 0.040'tır.

Oluşturulan diskriminant fonksiyonları, bağımlı değişkenlerdeki toplam varyansın ($0.677 + 0.040 = 0.717$) yaklaşık %71.7'sini açıklamaktadır. Bu sonuçlar, birinci fonksiyonun ayırt edici özelliğinin yüksek olduğunu göstermektedir.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.309	3514.188	6	<.001
2	.960	121.269	2	<.001

Şekil 56 Wilks Lambda Sonuçları

Tabloda (Şekil 56), ayırıcı fonksiyonların gruplar arasındaki farklılıkları açıklama gücüne ilişkin Wilks' Lambda testi sonuçları sunulmaktadır. Wilks' Lambda, modelin açıklayamadığı toplam varyansın bir göstergesi olup, daha düşük bir değer modelin daha başarılı olduğunu ifade eder.

"1 through 2" başlığı altındaki Wilks' Lambda değeri 0.309 olarak verilmiştir ve bu durum, modelin gruplar arasındaki farklılıkları açıklamada %69,1 oranında başarılı olduğunu, açıklayamadığı kısmın ise %30,9 olduğunu göstermektedir. İkinci fonksiyon için Wilks' Lambda değeri ise 0.960 olarak hesaplanmış ve modelin açıklayamadığı kısmın %96 olduğu ifade edilmiştir.

Chi-square testi ve anlamlılık (Sig.) değerleri incelendiğinde, her iki fonksiyon için de Chi-square değerlerinin yüksek olduğu ve anlamlılık düzeylerinin ($\text{Sig.} < 0,001$) 0,05'in altında olduğu görülmektedir. Bu sonuçlar doğrultusunda, sıfır hipotezi (H_0 : Ayırıcı fonksiyonların gruplar arasındaki farklılığı açıklama oranı anlamsızdır) reddedilmiş, alternatif hipotez (H_1 : Ayırıcı fonksiyonların gruplar arasındaki farklılığı açıklama oranı anlamlıdır) reddedilememiştir.

Sonuç olarak, elde edilen Wilks' Lambda değerleri ve anlamlılık testleri, ayırıcı fonksiyonların gruplar arasındaki farklılıkları anlamlı bir şekilde açıkladığını göstermektedir. Model istatistiksel olarak anlamlıdır ve gruplar arasındaki varyansın açıklanmasında etkili bir performans sergilemektedir.

Standardized Canonical Discriminant Function Coefficients		
	Function	
	1	2
daily_steps_transformed	-.386	.907
hydration_level_transformed	.042	.209
bmi_transformed	.930	.362

Şekil 57 Kanonik Ayırım Fonksiyonlarının Katsayıları

Tabloya göre (Şekil 57), Standartlaştırılmış Kanonik Diskriminant Fonksiyon Katsayıları, her bir değişkenin farklı fonksiyonlar üzerindeki göreceli etkisini göstermektedir. İlk fonksiyon için en büyük katkıyı sağlayan değişken, yüksek katsayısı ile "bmi_transformed" değişkeni olarak öne çıkmaktadır (0.930). İkinci fonksiyon için ise, en büyük katkıyı sağlayan değişken "daily_steps_transformed" olup 0.907 katsayı ile belirgin bir etkiye sahiptir.

$$F_1 = -0.386 \text{ daily steps transformed} + 0.042 \text{ hydration level transformed} + 0.930 \text{ bmi transformed}$$

$$F_2 = 0.907 \text{ daily steps transformed} + 0.209 \text{ hydration level transformed} + 0.362 \text{ bmi transformed}$$

Structure Matrix

	Function	
	1	2
bmi_transformed	.921 [*]	.370
calories_burned_transformed ^b	.157 [*]	.080
age_transformed ^b	.054 [*]	.015
avg_heart_rate_transformed ^b	-.045 [*]	-.017
duration_minutes_transformed ^b	-.036 [*]	-.019
blood_pressure_diastolic_transformed ^b	-.032 [*]	.005
daily_steps_transformed	-.369	.912 [*]
hydration_level_transformed	.005	.185 [*]
stress_level_transformed ^b	.000	.032 [*]
hours_sleep_transformed ^b	-.004	-.031 [*]
blood_pressure_systolic_transformed ^b	.020	.026 [*]
resting_heart_rate_transformed ^b	.000	.009 [*]

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

b. This variable not used in the analysis.

Şekil 58 Yapı Matrisi

Yukarıdaki Structure Matrix tablosuna göre (Şekil 58), fonksiyonlar ile ayrıştırıcı değişkenler arasındaki ilişkiyi gösteren korelasyonlar verilmiştir. Her bir değişkenin mutlak değer üzerinden sıralandığında, modele katkısı büyük olan değişkenin daha baskın olduğu görülmektedir.

Birinci fonksiyon için en yüksek mutlak korelasyon, bmi_transformed değişkeninde görülmektedir (0.921). Bu, birinci fonksiyonun bireylerin vücut kitle indeksi (BMI) ile güçlü bir şekilde ilişkilendirildiğini göstermektedir. İkinci fonksiyon için ise en büyük mutlak korelasyon daily_steps_transformed değişkeninde gözlenmiştir (0.912). Bu durum, ikinci fonksiyonun bireylerin günlük adım sayısı ile yüksek bir ilişki içinde olduğunu göstermektedir.

Diğer değişkenler arasında, birinci fonksiyonda calories_burned_transformed (0.157) ve ikinci fonksiyonda hydration_level_transformed (0.185) gibi değişkenler daha düşük bir korelasyona sahip olup, göreceli olarak daha az katkıda bulunmaktadır. Korelasyon değerlerine göre değişkenlerin katkıları incelendiğinde, hem birinci hem de ikinci fonksiyonda en büyük katkıyı sağlayan değişkenlerin sırasıyla bmi_transformed ve daily_steps_transformed olduğu söylenebilir.

Bu yapı matrisi, her bir fonksiyonun hangi değişkenlerle daha güçlü bir şekilde ilişkili olduğunu ve bu değişkenlerin modeldeki ayırt edici gücünü açık bir şekilde ortaya koymaktadır.

Prior Probabilities for Groups			
category	Prior	Cases Used in Analysis	
		Unweighted	Weighted
High	.333	999	999.000
Low	.333	1000	1000.000
Medium	.334	1001	1001.000
Total	1.000	3000	3000.000

Şekil 59 Gruplar için Nispi Olasılıklar

Şekil 59’da verilen prior olasılıkların kareleri toplamı, modelin doğruluk oranıyla kıyaslanarak modelin başarısı değerlendirilmektedir. Bu durumda, prior olasılıkların karelerinin toplamının doğruluk oranından küçük olması, modelin yeterli performansı sağladığını gösterir.

Prior olasılıkların kareleri şu şekilde hesaplanmıştır:

- $0,333=0,333^2 = 0,110889$
- $0,333=0,333^2 = 0,110889$
- $0,334=0,334^2 = 0,111556$
- Bu değerlerin toplamı: $0,110889 + 0,110889 + 0,111556 = 0,333334$

Modelin doğru atama oranı olan %76,4 (0,764) ile kıyaslandığında, prior değerlerin karelerinin toplamı (0,33) daha küçüktür. Bu durum, modelin doğruluk oranının prior olasılıkların

karelerinden daha büyük olduğunu ve modelin başarıyla sınıflandırma yaptığını göstermektedir.

Sonuç olarak, bu kıyaslama, modelin yeterli performansı sergilediğini ve başarılı bir sınıflandırma yöntemi olduğunu ortaya koymaktadır.

Functions at Group Centroids		
category	Function	
	1	2
High	1.836	-.128
Low	-1.709	-.159
Medium	-.124	.287

Unstandardized canonical discriminant functions evaluated at group means

Şekil 60 Grup Merkez Noktaları

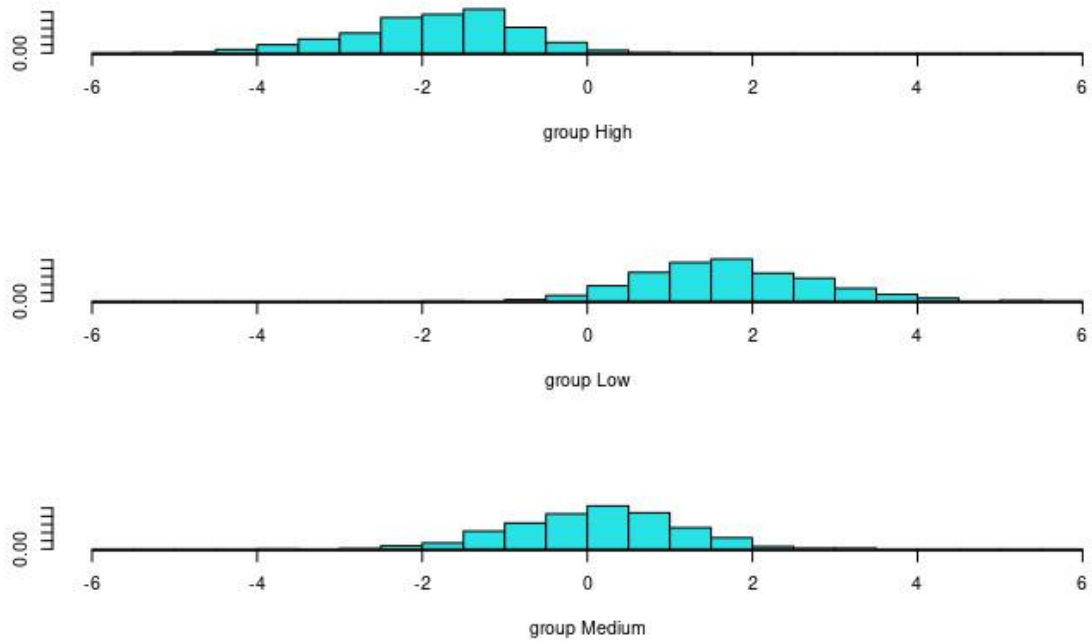
Yukarıdaki tablo (Şekil 60), her bir sınıfın kanonik diskriminant fonksiyonları üzerindeki skorlarını göstermektedir. Bu skorlar, modelin sonradan oluşturduğu değerlere dayanmaktadır ve doğrudan veri setinde mevcut değildir. Her bir skor, ilgili sınıfın diskriminant fonksiyonlarına göre yerleşimini ve sınıflar arası ayrışma derecesini temsil etmektedir.

Birinci fonksiyon incelendiğinde, "High" sınıfı için pozitif bir skor (1.836) elde edilirken, "Low" sınıfı için negatif bir skor (-1.709) görülmektedir. Bu iki sınıf arasındaki işaret farkı ve büyük skor farkı, birinci fonksiyonun bu iki sınıfı etkili bir şekilde ayırt edebildiğini göstermektedir. "Medium" sınıfının skoru (-0.124) ise diğer iki sınıfa göre daha orta bir konumda bulunmakta ve bu sınıfın birinci fonksiyondan etkilenme düzeyinin daha düşük olduğunu ortaya koymaktadır.

İkinci fonksiyon açısından değerlendirildiğinde, "Medium" sınıfının pozitif bir skor (0.287) aldığı, "High" (-0.128) ve "Low" (-0.159) sınıflarının ise negatif skorlar elde ettiği görülmektedir. Bu durum, ikinci fonksiyonun özellikle "Medium" sınıfını diğer iki sınıftan ayırmada daha başarılı olduğunu göstermektedir.

Sonuç olarak, sınıflar arasındaki işaret farklarının varlığı ve skorlar arasındaki belirgin farklılıklar, modelin sınıfları etkili bir şekilde ayırtılabildiğini ve başarılı bir sınıflandırma

performansı sergilediğini kanıtlamaktadır. Bu bulgular, modelin sınıflandırma doğruluğunu ve ayırıştırma kapasitesini desteklemektedir.



Şekil 61 Histogram Grafikleri

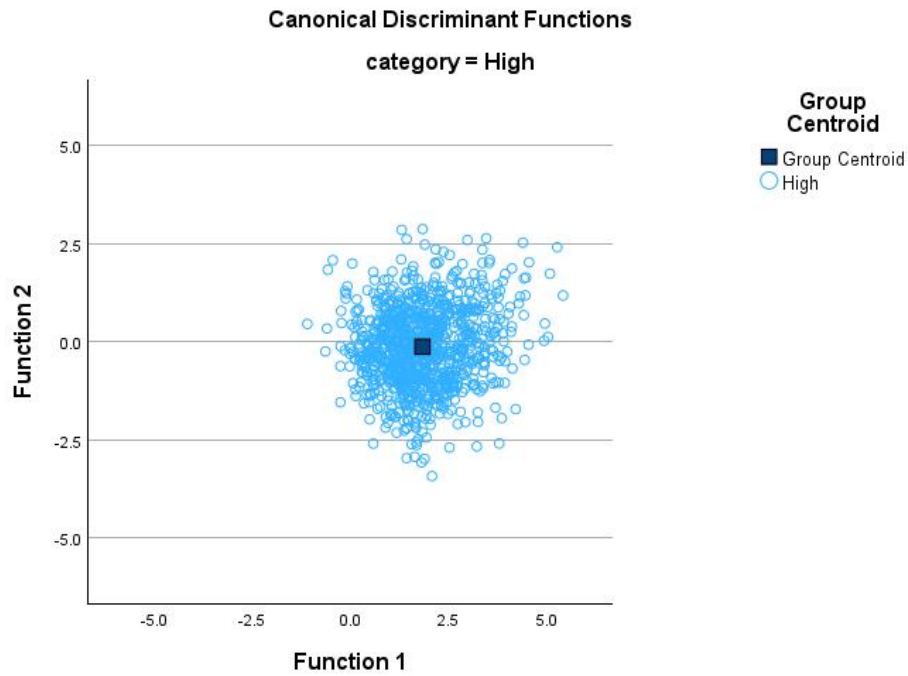
Yukarıdaki grafikler (Şekil 61), her bir sınıfın diskriminant skorlarının dağılımını göstermektedir. "High", "Low" ve "Medium" sınıflarına ait histogramlar, diskriminant fonksiyonları tarafından sınıfların nasıl ayırıştırıldığını görselleştirmektedir. Her bir sınıf için skorların farklı bir aralıkta yoğunlaşması, modelin sınıflar arasındaki ayrımı başarıyla gerçekleştirdiğine işaret etmektedir.

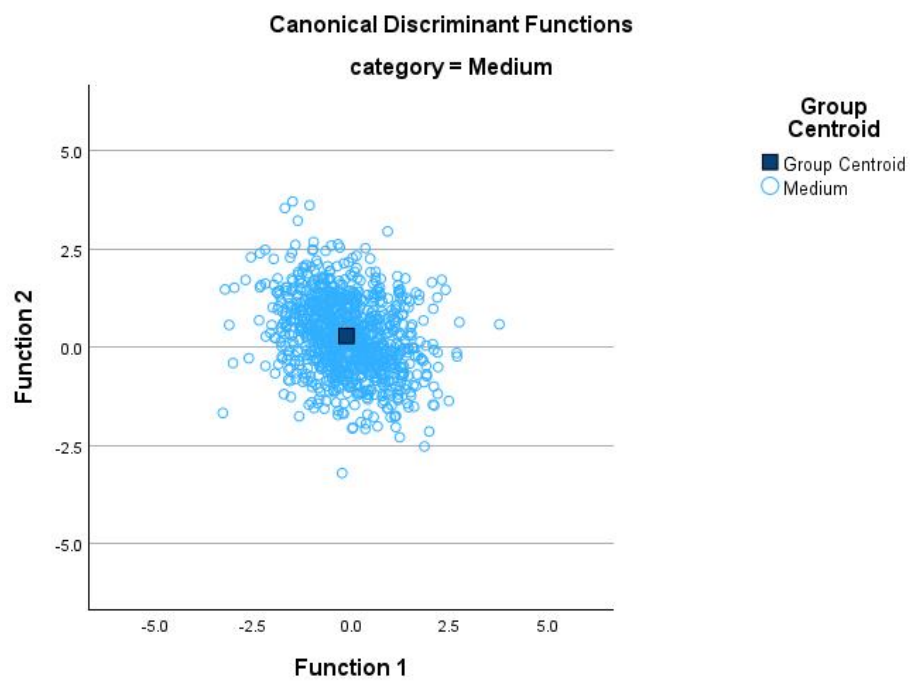
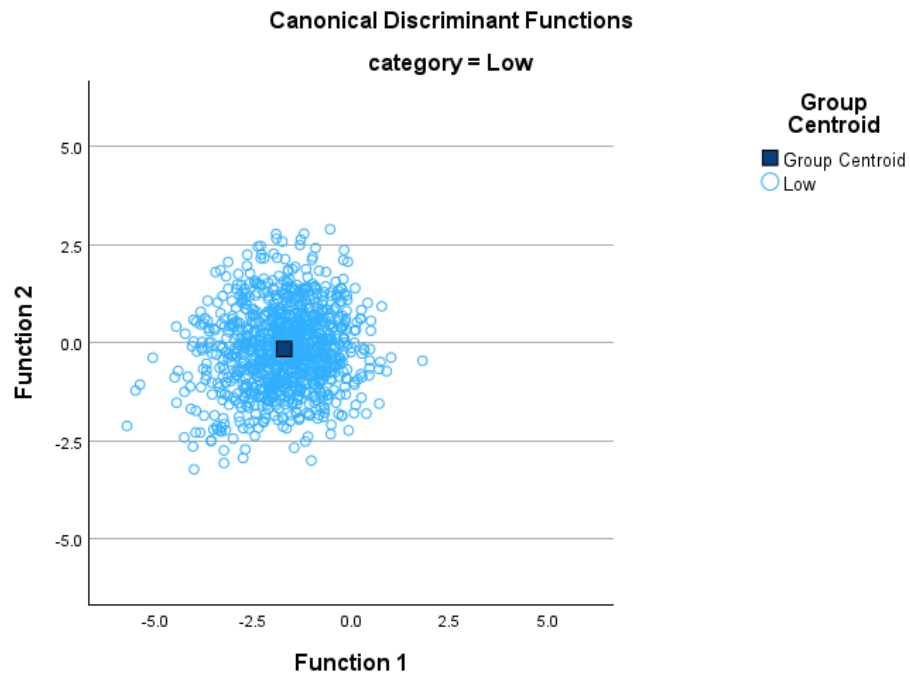
Birinci grup için, "High" grubu için diskriminant skorları ağırlıklı olarak negatif değerler üzerinde yoğunlaşmıştır ve bu grup için skorların dağılımı diğer gruplardan farklı bir şekilde şekillenmiştir. Bu, "High" sınıfının diskriminant fonksiyonları tarafından belirgin bir şekilde ayrıldığını göstermektedir.

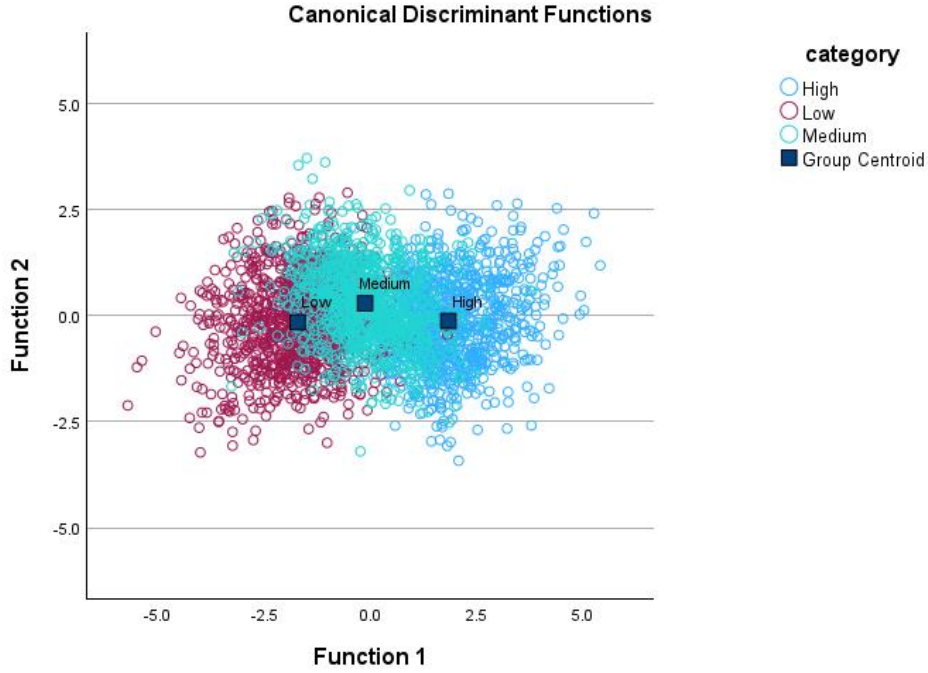
İkinci grup için, "Low" grubu için skorlar daha çok pozitif değerlere yakın bir bölgede yoğunlaşmaktadır. Bu grubun skorlarının dağılımı, "High" ve "Medium" sınıflarından ayrıştığını göstermektedir.

"Medium" grubu için skorların dağılımı, genel olarak pozitif değerlerde yoğunlaşsa da, diğer iki gruba göre daha simetrik bir dağılım sergilemektedir. Bu, "Medium" grubunun diğer iki sınıfa göre daha farklı bir yapıya sahip olduğunu ve diskriminant fonksiyonları tarafından etkili bir şekilde ayrıştırıldığını göstermektedir.

Grafikte, her bir grubun skor dağılımlarının farklı aralıklarda yoğunlaşması, sınıfların diskriminant fonksiyonları tarafından iyi bir şekilde ayrıştırıldığını kanıtlamaktadır. Modelin, sınıflar arasındaki farklılıkları etkili bir şekilde tespit ederek başarılı bir sınıflandırma gerçekleştirdiği söylenebilir.







Şekil 62 Centroid Grafikleri

Yukarıdaki grafik, kanonik diskriminant fonksiyonlarının birinci ve ikinci eksenler üzerindeki sınıfları nasıl ayırdığını görselleştirmektedir. "High," "Medium," ve "Low" kategorilerindeki örneklerin dağılımları ile birlikte grup centroid noktaları (siyah kareler) da gösterilmiştir.

Grafikte, her bir kategoriye ait veri noktalarının birbirinden farklı alanlarda yoğunlaştığı görülmektedir. Özellikle, fonksiyon 1 eksenini boyunca "High" ve "Low" gruplarının belirgin bir şekilde ayrıldığı dikkat çekmektedir. "Medium" grubu ise bu iki grup arasında, daha orta bir konumda yer almakta ve iki fonksiyonun da kombinasyonu ile ayrılmaktadır. Bu durum, fonksiyon 1'in sınıflar arasında en güçlü ayırıcı fonksiyon olduğunu göstermektedir.

Fonksiyon 2 ekseninde ise, "Medium" grubunun diğer gruplardan farklılaşmasında etkili olduğu görülmektedir. Bu fonksiyon, özellikle "Medium" grubunun diğer iki grup ile ayrışmasını güçlendiren bir ayrım yaratmaktadır.

Her ne kadar bazı noktaların kesişmesi doğal bir durum olarak kabul edilse de, bu grafik için kesişme bölgelerinin oldukça sınırlı olduğu görülmektedir. Bu durum, sınıflar arası farkların diskriminant fonksiyonları ile etkili bir şekilde yakalandığını ve modelin, rating türlerini ayırt etme açısından güçlü bir performans sergilediğini açıkça ortaya koymaktadır. Bu nedenle, hem sınıflar arasındaki ayrım hem de kesişme bölgelerinin minimal düzeyde olması, modelin genel başarısını desteklemektedir.

Classification Results^a

		Predicted Group Membership				
		category	High	Low	Medium	Total
Original	Count	High	848	1	150	999
		Low	6	778	216	1000
		Medium	176	160	665	1001
	%	High	84.9	.1	15.0	100.0
		Low	.6	77.8	21.6	100.0
		Medium	17.6	16.0	66.4	100.0

a. 76.4% of original grouped cases correctly classified.

Şekil 63 Sınıflandırma Sonuçları

Şekil 63'teki sınıflandırma sonuçları, modelin her bir sınıf için doğru ve yanlış sınıflandırma oranlarını göstermektedir. Verilen bilgiler, modelin genel performansını ve sınıflar bazındaki başarısını detaylandırmaktadır.

Birinci grup (High) için toplamda 999 örnek bulunmaktadır ve bu örneklerin 848'i doğru şekilde sınıflandırılmıştır. Bu, birinci grup için doğru sınıflandırma oranının **%84,9** olduğunu göstermektedir. Yanlış sınıflandırılan örneklerin büyük kısmı "Medium" grubuna atanmıştır.

İkinci grup (Low) için 1000 örnek yer almakta ve bu örneklerden 778'i doğru şekilde sınıflandırılmıştır. Doğru sınıflandırma oranı **%77,8**'dir. Yanlış sınıflandırılan örneklerin bir kısmı "Medium" grubuna atanmıştır.

Üçüncü grup (Medium) için toplamda 1001 örnek bulunmaktadır ve bunların 665'i doğru sınıflandırılmıştır. Bu grubun doğru sınıflandırma oranı **%66,4** olarak hesaplanmıştır. Yanlış sınıflandırılan örneklerin çoğunluğu "High" ve "Low" gruplarına atanmıştır.

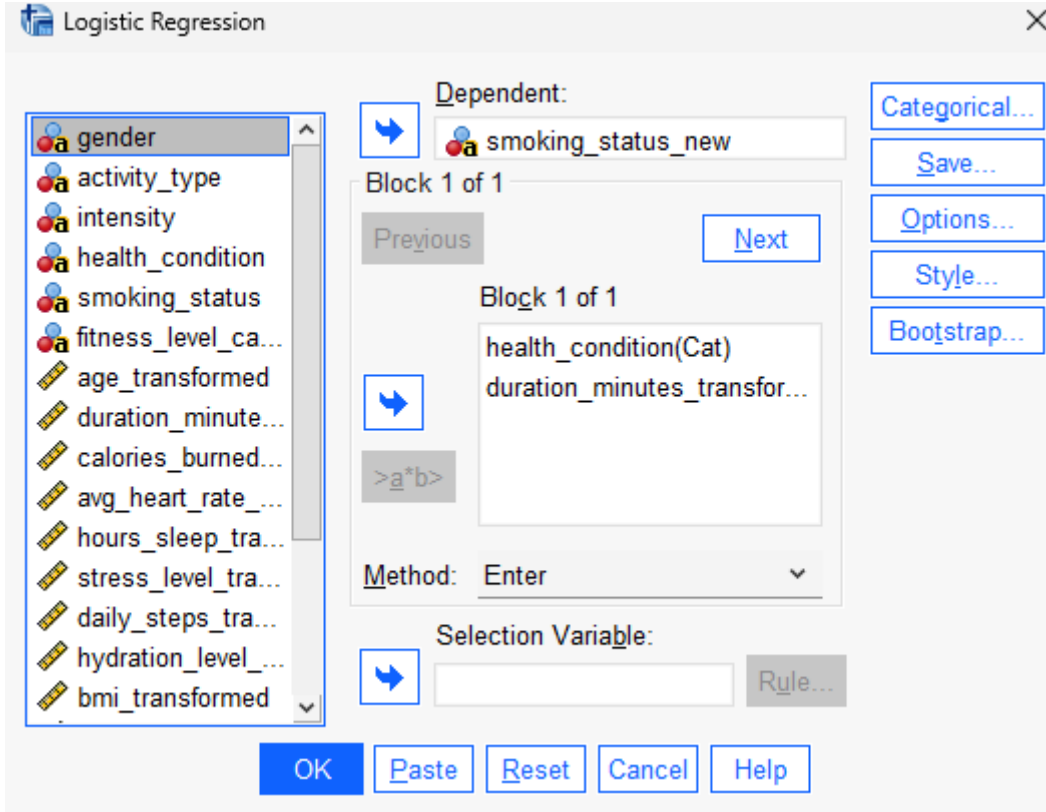
Tablonun alt kısmında belirtildiği gibi, model toplamda tüm verilerin **%76,4'ünü** doğru şekilde sınıflandırmıştır. Bu sonuçlar, modelin genel olarak iyi bir sınıflandırma performansı sergilediğini, ancak bazı sınıflar arasında kısmi bir karışıklık olduğunu göstermektedir.

8. Lojistik Regresyon Analizi

Lojistik regresyon, kategorik bir bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılan bir istatistiksel yöntemdir. Özellikle bağımlı değişkenin iki sınıfa ayrıldığı durumlarda ikili lojistik regresyon, daha fazla sınıfa ayrıldığı durumlarda ise çoklu lojistik regresyon uygulanmaktadır. Bu yöntem, bağımlı değişkenin bir sınıfa ait olma olasılığını, bağımsız değişkenlerin bir fonksiyonu olarak tahmin eder ve tahminler genellikle logit (logaritmik odds) dönüşümü üzerinden gerçekleştirilir.

Lojistik regresyon, tıp, ekonomi, sosyal bilimler ve mühendislik gibi birçok alanda yaygın bir şekilde kullanılmaktadır. Örneğin, bir hastanın belirli bir hastalığa sahip olup olmadığını tahmin etmek, müşterilerin bir ürünü satın alma olasılığını modellemek veya bireylerin bir karar verme sürecindeki eğilimlerini analiz etmek gibi pek çok uygulama alanına sahiptir. Bu yöntem, bağımsız değişkenler arasında doğrusal olmayan ilişkileri ele alabilmesi ve sürekli veya kategorik bağımsız değişkenlerle çalışabilmesi nedeniyle esnek ve güçlü bir analiz aracı olarak kabul edilmektedir.

Bu çalışmada lojistik regresyon yöntemi, bağımlı değişkenin kategorik yapısı dikkate alınarak uygulanacaktır. Özellikle fitness seviyeleri gibi kategorik verilerin analizinde kullanılacak olan bu yöntem, bağımsız değişkenler ile fitness seviyeleri arasındaki ilişkileri anlamada ve modellemede önemli bir rol oynayacaktır. Lojistik regresyonun sağladığı olasılık tahminleri, ilgili değişkenlerin etkilerini yorumlama ve çıkarım yapma açısından değerli bilgiler sunacaktır.



Bu çalışma, spor yapan bireylerin sigara içme durumu ile sağlık durumları ve egzersiz süreleri arasındaki ilişkiyi incelemeyi amaçlamaktadır. Araştırma, sigara içme alışkanlığına sahip bireylerin sağlık profillerinin ve fiziksel aktivite sürelerinin bu davranış üzerindeki etkisini anlamayı hedeflemektedir.

Analizde bağımlı değişken olarak sigara içme durumu, bağımsız değişkenler olarak ise sağlık durumu ve egzersiz süreleri kullanılmıştır. Sağlık durumu değişkeni, bireylerin genel sağlık profillerini yansıtan kategorik bir değişken olarak modele dahil edilmiştir (örneğin, sağlıklı, hipertansiyon, diyabet). Egzersiz süresi değişkeni ise sürekli bir değişken olarak değerlendirilmiş ve bireylerin fiziksel aktivite süreleri ile sigara içme durumu arasındaki ilişki analiz edilmiştir.

Bu analizde lojistik regresyon modeli uygulanmıştır. Model, sağlık durumu ve egzersiz sürelerinin bireylerin sigara içme durumunu ne ölçüde etkilediğini ve bu etkinin istatistiksel olarak anlamlı olup olmadığını ortaya koymaktadır.

Classification Table^{a,b}

Observed			Predicted		Percentage Correct
			smoking_status_new no	yes	
Step 0	smoking_status_new	no	1818	0	100.0
		yes	1182	0	.0
	Overall Percentage				

a. Constant is included in the model.

b. The cut value is .500

Şekil 64 Sınıflandırma Tablosu

Tahmin edilme sürecinde, bağımlı değişken iki kategoriye atanmıştır. Rastgele sabit terim içeren model kullanılarak gerçekleştirilen sınıflandırmada, başarılı sınıflama oranı %60,6 olarak elde edilmiştir. (Şekil 64)

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	health_condition	3.271	3	.352
		health_condition(1)	.691	1	.406
		health_condition(2)	.090	1	.765
		health_condition(3)	.406	1	.524
		duration_minutes_transformed	.347	1	.556
		Overall Statistics	3.714	4	.446

Şekil 65 Eşitlikte Bulunmayan Değişkenler

Modele eklenen bağımsız değişkenlerin katkıları, tabloya göre (Şekil 65) skor değerleri üzerinden değerlendirilmiştir. Tabloda en yüksek score değerine sahip değişkenin health_condition olduğu görülmektedir (3.271). Ancak, bu değişkenin anlamlı bir katkı sağlamadığı, Sig. değeri (0.352) ile belirlenmiştir. duration_minutes_transformed değişkeni de modele katkı sağlamamış olup, score değeri (0.347) ve anlamlılık seviyesi (0.556) ile bu durumu desteklemektedir.

Tablodaki tüm bağımsız değişkenlerin anlamlılık seviyeleri (Sig.) 0.05'ten büyük olduğu için, modelde anlamlı katkı sağlayan bir değişken bulunmamaktadır. Ayrıca, Overall Statistics

değeri (Sig. = 0.446) de tüm değişkenlerin topluca modele anlamlı bir katkı sunmadığını göstermektedir.

Sonuç olarak, modele dahil edilen bağımsız değişkenlerin, hem bireysel hem de toplu olarak, istatistiksel açıdan anlamlı bir etkisi olmadığı söylenebilir. Bu durum, modelin açıklayıcılığını artırmak için farklı bağımsız değişkenlerin dahil edilmesi gerektiğini gösterebilir.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4019.336 ^a	.001	.002

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Şekil 66 Model Özeti

Tabloda sunulan Model Summary (Şekil 66), modelin genel uygunluğunu ve açıklayıcılığını değerlendirmek için kullanılan önemli istatistiksel ölçütleri içermektedir. -2 Log Likelihood değeri, modelin uyumunu temsil eder ve bu örnekte 4019.336 olarak hesaplanmıştır. Bu değerin yüksek olması, modelin veri setine uygunluğunun sınırlı olabileceğini göstermektedir.

Cox & Snell R Square ve Nagelkerke R Square değerleri, modelin açıklayıcılık gücünü yansıtmaktadır. Cox & Snell R Square değeri 0.001, Nagelkerke R Square değeri ise 0.002 olarak hesaplanmıştır. Bu değerler oldukça düşük olup, modelin bağımlı değişken üzerindeki toplam varyansın çok küçük bir kısmını açıkladığını göstermektedir. Literatürde bu değerlerin genellikle 0.30 veya daha yüksek olmasının istenildiği belirtilmekte olup, düşük R kare değerleri modelin açıklayıcılığının zayıf olduğunu işaret eder.

R kare değerlerinin regresyon analizlerine kıyasla daha düşük çıkma eğiliminde olduğu bilinmesine rağmen, bu kadar düşük değerler modelin veri setini yeterince açıklayamadığını göstermektedir. Ancak, R kare değerlerinin düşük olmasının modelin tamamen geçersiz olduğu anlamına gelmediği ve bu değerlere ek olarak diğer model uygunluk kriterlerinin de dikkate alınması gerektiği unutulmamalıdır.

Sonuç olarak, -2 Log Likelihood değerinin yüksek olması ve R kare değerlerinin düşük seviyelerde kalması, kullanılan veri setinin model için uygun olmayabileceğini ve modelin bağımlı değişkeni açıklama gücünün sınırlı olduğunu düşündürmektedir. Bu durum, modelin

performansını artırmak için veri setinde veya kullanılan değişkenlerde revizyon yapılması gerektiğini işaret edebilir.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	19.983	8	.010

Şekil 67 Hosmer and Lemeshow Test Sonuçları

Hosmer ve Lemeshow Testi sonuçlarına göre (Şekil 67), modelin uyumunu değerlendiren hipotezler şu şekildedir:

- **H0:** Beklenen olaylar ile gözlemlenen olaylar arasında anlamlı bir fark yoktur, dolayısıyla model veri ile uyumludur.
- **H1:** Beklenen olaylar ile gözlemlenen olaylar arasında anlamlı bir fark vardır, dolayısıyla model veri ile uyumlu değildir.

Tabloda, Hosmer ve Lemeshow testi için Chi-square değeri 19.983, serbestlik derecesi (df) 8, ve anlamlılık değeri (Sig.) 0.010 olarak hesaplanmıştır. Sig. değeri 0.05'ten küçük olduğu için, H0 hipotezi reddedilir. Bu durumda, beklenen ve gözlemlenen olaylar arasında anlamlı bir fark olduğu söylenebilir.

Bu sonuç, modelin veri ile uyumlu olmadığını ve modelin gözlemlenen olayları yeterince iyi açıklamadığını göstermektedir. Hosmer ve Lemeshow testi, modelin genel uygunluğunu değerlendirmek için önemli bir ölçüttür, bu nedenle elde edilen bu sonuç modelin performansının iyileştirilmesi gerektiğine işaret etmektedir. Daha iyi bir model uyumu için ek değişkenler eklenebilir veya mevcut değişkenlerin etkileşimleri dikkate alınabilir.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	health_condition			3.357	3	.340	
	health_condition(1)	-.353	.207	2.893	1	.089	.703
	health_condition(2)	-.249	.173	2.080	1	.149	.780
	health_condition(3)	-.316	.194	2.645	1	.104	.729
	duration_minutes_transformed	-.025	.038	.444	1	.505	.975
	Constant	-.175	.167	1.093	1	.296	.840

a. Variable(s) entered on step 1: health_condition, duration_minutes_transformed.

Şekil 68 Eşitlikteki Değişkenler

Tabloya göre (Şekil 68), Wald testi istatistikleri, her bir bağımsız değişkenin modele kattığı anlamlılığı değerlendirmek için kullanılmıştır. Wald test istatistiği ne kadar büyükse, o değişkenin modele anlamlı katkısı o kadar yüksek olur. Verilen değerler ışığında, Wald istatistiklerine göre bağımsız değişkenlerin modele kattığı anlamlılık sıralaması şu şekildedir:

- health_condition (1): 2.893
- health_condition (3): 2.645
- health_condition (2): 2.080
- duration_minutes_transformed: 0.444

Bu sıralamaya göre, health_condition (1) değişkeni modele en yüksek anlamlı katkıyı sunarken, diğer değişkenlerin katkıları daha düşüktür. Ancak, tüm değişkenler için Sig. değerleri 0.05'ten büyük olduğundan, modele anlamlı katkı sağlayan bir değişken bulunmamaktadır.

Regresyon Denklemi: Modelin regresyon denklemi şu şekilde ifade edilebilir:

$$p = -0.175 - 0.353 * \text{health_condition}(1) - 0.249 * \text{health_condition}(2) - 0.316 * \text{health_condition}(3) - 0.025 * \text{duration_minutes_transformed}$$

Burada **B** katsayıları, her bir bağımsız değişkenin logaritmik odds üzerindeki etkisini temsil eder.

- **health_condition(1):** Sağlık durumu kategorisinin 1 olması, bağımlı değişkenin olasılığını **0.703** oranında azaltır.

- **health_condition(2):** Sağlık durumu kategorisinin 2 olması, bağımlı değişkenin olasılığını **0.780** oranında azaltır.
- **health_condition(3):** Sağlık durumu kategorisinin 3 olması, bağımlı değişkenin olasılığını **0.729** oranında azaltır.
- **duration_minutes_transformed:** Egzersiz süresindeki bir birimlik artış, bağımlı değişkenin olasılığını **0.975** oranında azaltır.

Sonuç olarak, tüm Sig. değerlerinin 0.05'ten büyük olması, modelde bağımsız değişkenlerin istatistiksel olarak anlamlı bir katkı sağlamadığını göstermektedir. Bu, modelin açıklayıcılığını artırmak için ek değişkenlerin dahil edilmesi veya mevcut değişkenlerin yeniden düzenlenmesi gerektiğine işaret eder.

9. Kümeleme Analizi

Kümeleme analizi, veri setinde bulunan tüm değişkenleri dikkate alarak, gözlemleri benzer özelliklere sahip gruplara ayırmayı amaçlayan bir istatistiksel yöntemdir. Bu projede, spor yapan bireylerin çeşitli sağlık ve yaşam tarzı değişkenlerine göre gruplandırılması hedeflenmiştir. Kullanılan değişkenler arasında sigara içme durumu (**smoking_status_new**), sağlık durumu (**health_condition**), egzersiz süreleri (**duration_minutes_transformed**), kalori yakımı, stres seviyeleri ve diğer fiziksel aktivite göstergeleri yer almaktadır. Bu kapsamlı değişken seti, bireyler arasında gizli kalmış kalıpları ve ilişkileri keşfetmek için güçlü bir temel sunmaktadır.

Kümeleme analizi, bağımlı değişken olmaksızın, veri setindeki benzerliklere dayanarak gruplar oluşturur. Bu yöntem, bireylerin genel sağlık profillerini, egzersiz alışkanlıklarını ve yaşam tarzlarını dikkate alarak heterojen bir veri setini daha anlamlı alt gruplara ayırmayı sağlar. Bu gruplar, belirli sağlık durumları, fiziksel aktivite seviyeleri ve davranışsal özellikler açısından homojen alt kümeler oluşturmayı hedefler.

Projede, kümeleme analizi kullanılarak spor yapan bireyler, sağlık durumları ve fiziksel aktivitelerine göre gruplanmıştır. Bu gruplama, sağlık profillerinin iyileştirilmesi, fiziksel aktivite alışkanlıklarının anlaşılması ve bireylerin genel yaşam kalitelerinin artırılmasına yönelik stratejiler geliştirilmesi açısından değerli bilgiler sunmaktadır. Analiz sonuçları, bireyler arasında ortak özelliklerin belirlenmesine, farklı grupların ihtiyaçlarının anlaşılmasına ve bu gruplara yönelik özel yaklaşımların oluşturulmasına katkı sağlayacaktır.

Kümeleme analizi için yöntem seçimi sırasında **Between-groups** ve **Ward's yöntemi** karşılaştırılmış ve en uygun yöntemin **Ward's yöntemi** olduğu belirlenmiştir. Bu yöntemin, gruplar arasındaki toplam varyansı minimize etmesi ve homojen gruplar oluşturması nedeniyle tercih edildiği gözlemlenmiştir. Ayrıca, uzaklık ölçütü olarak **Öklid uzaklığı** kullanılmış ve veri setinin yapısı ile analiz verimliliği dikkate alınarak bu yaklaşımın daha uygun olduğu sonucuna varılmıştır. Bu seçimlerin, veri setindeki gözlemlerin anlamlı gruplar halinde sınıflandırılmasını kolaylaştırdığı düşünülmektedir.

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	287	730	.025	0	0	1906
2	906	2543	.054	0	0	1848
3	1091	1160	.090	0	0	13
4	1591	2902	.127	0	0	50
5	1324	2458	.165	0	0	854
6	2038	2377	.204	0	0	1482
7	955	2114	.245	0	0	355
8	312	1111	.287	0	0	1790
9	1844	2487	.331	0	0	1069
10	458	1023	.376	0	0	1143
11	879	2585	.422	0	0	2008
12	1022	2539	.469	0	0	568
13	932	1091	.517	0	3	182
14	210	2450	.566	0	0	367
15	337	1841	.615	0	0	1628
16	279	1963	.665	0	0	1006
17	299	2259	.715	0	0	516
18	953	2644	.766	0	0	230
19	74	2777	.816	0	0	1574
20	1571	2977	.867	0	0	1868

Şekil 69 Kümelenme Düzeni

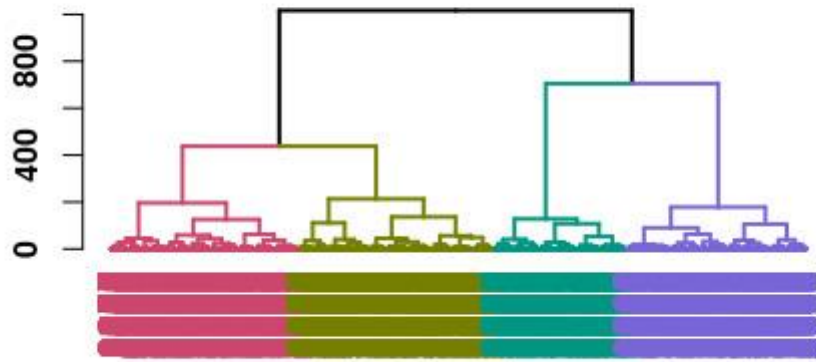
Tablodaki Agglomeration Schedule (Şekil 69) verileri incelendiğinde, her bir adımda birleşen kümeler ve bu birleşmelerin gerçekleştiği uzaklık değerleri gözlemlenmiştir. Örnek olarak:

- **287. gözlem** ile **730. gözlem**, **0.025** uzaklıkta birleşmiştir.
- **906. gözlem** ile **2543. gözlem**, **0.054** uzaklıkta birleşmiştir.
- **1091. gözlem** ile **1160. gözlem**, **0.090** uzaklıkta birleşmiştir.
- **1591. gözlem** ile **2902. gözlem**, **0.127** uzaklıkta birleşmiştir.

Bu aşamalar, her iki kümenin birleşme noktalarında ölçülen en küçük uzaklığı temsil etmektedir. Uzaklık değeri arttıkça, birleşen kümelerin birbirine olan benzerliklerinin azaldığı ve daha uzak kümelerin birleştirildiği gözlemlenmektedir. Bu süreçteki örnekler, kümeleme algoritmasının verileri nasıl adım adım homojen gruplar oluşturacak şekilde birleştirdiğini göstermektedir.

Ayrıca, uzaklık değerleri büyüdükçe daha büyük grupların birleştiği anlaşılmaktadır. Örneğin:

- **312. gözlem** ile **1111. gözlem**, **0.287** uzaklıkta birleşmiştir.
- **210. gözlem** ile **2450. gözlem**, **0.566** uzaklıkta birleşmiştir.
- **1571. gözlem** ile **2977. gözlem**, **0.867** uzaklıkta birleşmiştir.



Şekil 70 Dendrogram

Yukarıdaki dendrogram incelendiğinde, gözlemlerde belirgin bir aykırı değer görülmemektedir. Kümeleme yapısına göre, dendrogramda farklı grupların net bir şekilde ayrıldığı ve verinin homojen bir şekilde dağıldığı anlaşılmaktadır. Kesme noktası dikkate alındığında, küme sayısı 4 olarak belirlenebilir. Bu durumda, veri seti, birbirinden farklı özelliklere sahip dört ana gruba ayrılmaktadır. Bu ayırım, gözlemlerin birbirine olan benzerlik düzeylerini dikkate alarak anlamlı gruplar oluşturduğunu göstermektedir.

Initial Cluster Centers				
	Cluster			
	1	2	3	4
duration_minutes_transformed	-1.495950071	-1.918876226	1.2489956451	1.5981931399
calories_burned_transformed	-1.433669266	-2.624930456	1.6901461375	.40701087626
daily_steps_transformed	3.5879146723	-1.903310819	.54997964728	-.7186877491
bmi_transformed	.07191862645	1.1926665268	3.5879146723	-2.465907383

Şekil 71 Başlangıç Küme Merkezleri

Tablodaki başlangıç küme merkezleri (Şekil 71), dört küme için rastgele seçilmiş değerleri temsil etmektedir ve bu merkezler, algoritmanın ilk iterasyonunda gözlemlerin kümelere atanmasını belirlemek için kullanılır. **Cluster 1**, düşük egzersiz süresi ve kalori yakımına rağmen yüksek adım sayısı ile dikkat çekmektedir ve BMI değerleri ortalamaya yakın bir düzeydedir. **Cluster 2**, düşük egzersiz süresi, en düşük kalori yakımı ve düşük adım sayısına sahip olup, BMI değerleri nispeten yüksektir. **Cluster 3**, yüksek egzersiz süresi ve kalori yakımı ile öne çıkmakta, ayrıca en yüksek BMI değerine sahiptir. **Cluster 4**, en yüksek egzersiz süresi ve orta düzeyde kalori yakımını gösterirken, düşük adım sayısı ve en düşük BMI değerleriyle dikkat çekmektedir.

Iteration History ^a				
Change in Cluster Centers				
Iteration	1	2	3	4
1	2.528	2.455	2.654	2.350
2	.349	.089	.344	.214
3	.199	.087	.101	.171
4	.145	.092	.041	.105
5	.099	.084	.026	.050
6	.063	.056	.030	.037
7	.035	.040	.030	.025
8	.027	.045	.032	.013
9	.030	.043	.029	.015
10	.025	.033	.029	.016
11	.024	.017	.011	.016
12	.014	.016	.016	.013
13	.010	.016	.019	.016
14	.011	.014	.013	.010
15	.005	.008	.008	.005
16	.002	.007	.011	.004
17	.007	.004	.011	.011
18	.003	.002	.006	.006
19	.003	.000	.004	.005
20	.000	.002	.000	.002
21	.000	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 21. The minimum distance between initial centers is 5.745.

Şekil 72 Kümeleme Geçmişi

72. tabloda, kümeleme algoritmasının her iterasyonda küme merkezlerindeki değişimi gösteren bir iterasyon geçmişi sunulmaktadır. İlk iterasyonda küme merkezlerinde büyük değişiklikler gözlenmiş olup, bu değişiklikler sonraki iterasyonlarda giderek azalmıştır. Algoritma, küme merkezlerindeki değişikliklerin önemsiz hale geldiği ve istikrar kazandığı **21. iterasyonda** sonlanmıştır. Bu noktada, algoritma **küme merkezlerinin final değerlerine** ulaşmıştır.

Tablonun altındaki açıklamada belirtildiği gibi, maksimum mutlak koordinat değişiminin **0.000'a** ulaşmasıyla algoritma yakınsama sağlamış ve işlem tamamlanmıştır. Bu, kümeler arasındaki mesafelerin ve merkezlerin sabit hale geldiğini ve algoritmanın optimal çözümü bulunduğunu göstermektedir. Minimum başlangıç küme merkezleri arasındaki mesafenin **5.745**

olması, kümeler arasındaki farklılıkların belirgin olduğunu ve algoritmanın iyi bir ayrım yaptığını desteklemektedir.

Final Cluster Centers				
	Cluster			
	1	2	3	4
duration_minutes_transformed	-.8980753315	-.7346819285	.80519887617	.60684202082
calories_burned_transformed	-1.081342232	-.4197106645	1.0827372384	.29393629601
daily_steps_transformed	.56616868214	-.7583954037	-.5714139525	.58149659382
bmi_transformed	-.7097896602	.76050885384	.80820790972	-.6594349206

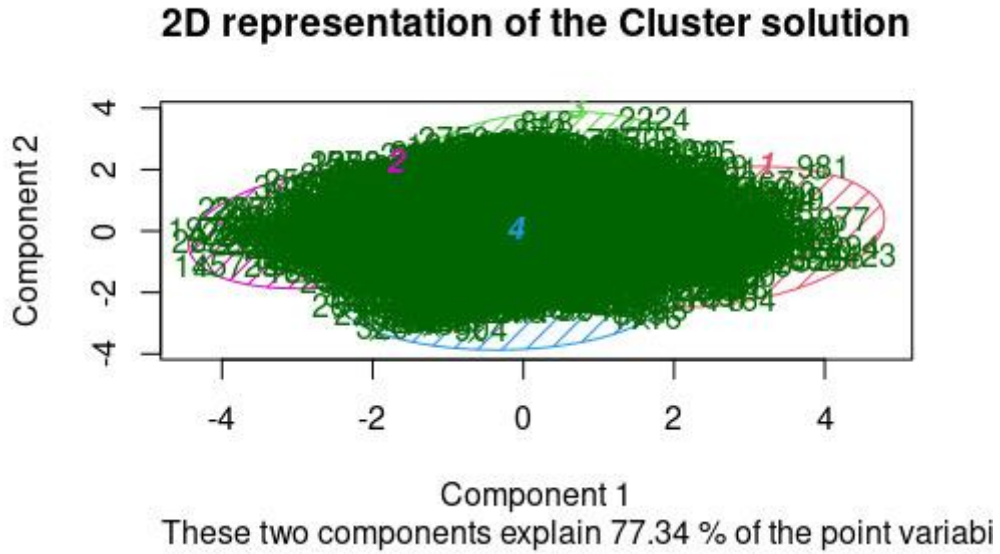
Şekil 73 Final Küme Merkezleri

Final küme merkezleri, kümeleme algoritmasının sonucunda elde edilen kümelerin ortalama değerlerini temsil etmekte ve her bir kümenin karakteristik özelliklerini ortaya koymaktadır. **1. Küme**, "duration_minutes_transformed" ve "calories_burned_transformed" değişkenlerinde en düşük ortalamalara sahip olup, genellikle düşük fiziksel aktivite süresi ve kalori yakımı ile karakterize edilmektedir. Bununla birlikte, "daily_steps_transformed" değişkeninde pozitif bir ortalamaya sahiptir. **2. Küme**, "bmi_transformed" değişkeninde en yüksek ortalamaya sahip olup, daha yüksek vücut kitle indeksine sahip bireyleri temsil etmektedir. Aynı zamanda "calories_burned_transformed" ve "daily_steps_transformed" değişkenlerinde negatif değerler ile düşük aktivite ve kalori yakımı göstermektedir. **3. Küme**, "duration_minutes_transformed" ve "calories_burned_transformed" değişkenlerinde en yüksek ortalamalara sahip olup, en yüksek fiziksel aktivite süresi ve kalori yakımı ile öne çıkmaktadır. Ancak, "daily_steps_transformed" değişkeninde negatif bir değere sahip olması, bu kümenin diğerlerine kıyasla daha düşük adım sayısına sahip olduğunu göstermektedir. **4. Küme**, "daily_steps_transformed" ve "duration_minutes_transformed" değişkenlerinde ortalama değerlere sahip olup, dengeli fiziksel aktivite ve adım sayısı ile karakterizedir. Bununla birlikte, "calories_burned_transformed" ve "bmi_transformed" değişkenlerinde orta düzeyde değerlere sahiptir.

Distances between Final Cluster Centers				
Cluster	1	2	3	4
1		2.093	3.344	2.039
2	2.093		2.160	2.474
3	3.344	2.160		2.036
4	2.039	2.474	2.036	

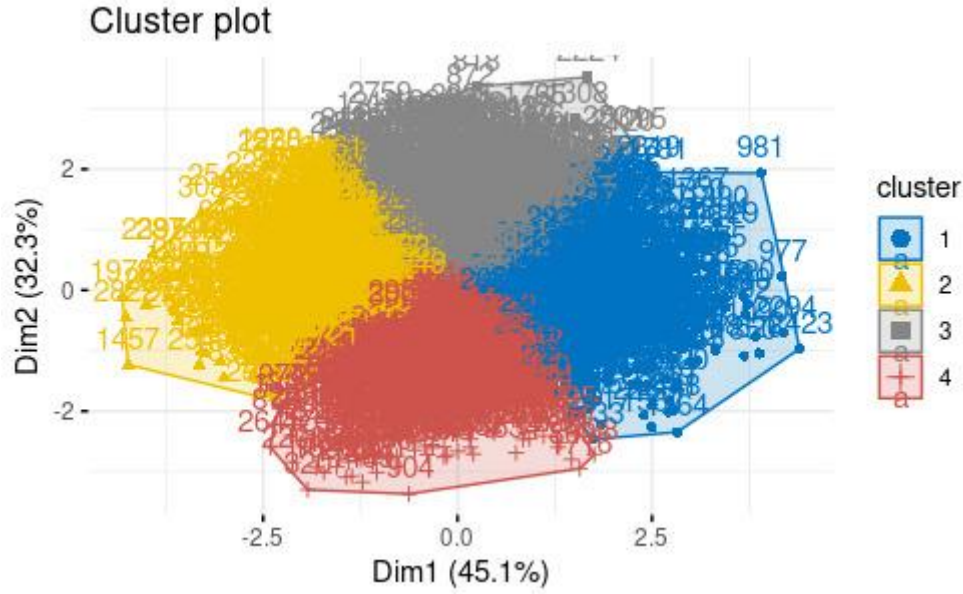
Şekil 74 Küme Merkezleri Arası Uzaklıklar

Tablodaki değerler (Şekil 74), her bir kümenin diğer kümelere olan uzaklıklarını temsil etmektedir. Bu uzaklıklar, kümelerin birbirine olan benzerlik ya da farklılıklarını ölçmek için kullanılır. **1. küme**, **2. küme**ye 2.093 birim uzaklıkta iken, **3. küme**ye olan uzaklığı 3.344 birimdir. Bu, **1. küme** ile **3. küme** arasında daha fazla farklılık olduğunu göstermektedir. En uzak kümeler ise, **2. küme** ile **3. küme** arasında olup, bu iki kümenin uzaklığı 2.160 birimdir.



Şekil 75 2 Boyutlu Kümeleme Gösterimi

Yukarıdaki grafik, kümeleme analizinin iki boyutlu bir temsilini sunmaktadır ve farklı kümelerin nasıl ayrıldığını görselleştirmektedir. Grafik üzerinde, iki bileşen toplamda verinin %77.34'ünü açıklamaktadır. Bu oran oldukça yüksek olup, kümeleme analizinin genel olarak başarılı bir şekilde gerçekleştirilmiş olduğunu göstermektedir.



Şekil 76 Kümeleme Gösterimi

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
duration_minutes_transformed	576.312	3	.409	2996	1408.032	<.001
calories_burned_transformed	619.227	3	.380	2996	1627.595	<.001
daily_steps_transformed	384.738	3	.616	2996	624.937	<.001
bmi_transformed	535.537	3	.465	2996	1152.591	<.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Şekil 77 ANOVA Tablosu

Yukarıdaki ANOVA tablosu, kümelerin dört temel değişken (**duration_minutes_transformed**, **calories_burned_transformed**, **daily_steps_transformed**, **bmi_transformed**) açısından ortalamaları arasında anlamlı farklar olup olmadığını incelemektedir. Bu bağlamda hipotezler şu şekilde tanımlanabilir:

- **H0:** Kümelerin ortalamaları arasında, incelenen değişkenler (egzersiz süresi, yakılan kalori, günlük adım sayısı, BMI) açısından anlamlı bir fark yoktur.

- **H1:** Kümelerin ortalamaları arasında, incelenen değişkenler açısından anlamlı farklar vardır.

Tabloda tüm değişkenler için **Sig.** değerleri < 0.001 olduğu için, **H0 hipotezi reddedilir** ve **H1 hipotezi reddedilemez**. Bu durum, kümelerin söz konusu değişkenler bakımından ortalamalar arasında istatistiksel olarak anlamlı farklılıklar gösterdiğini ortaya koymaktadır.

Ayrıca, **F** değerleri oldukça yüksektir (örneğin, "duration_minutes_transformed" için $F = 1408.032$), bu da değişkenler arasında kümeleri ayıran güçlü bir etkileyici faktör olduğunu göstermektedir. Bu sonuçlar, kümelerin belirgin bir şekilde farklı sağlık ve fiziksel aktivite profillerine sahip olduğunu ve analiz edilen değişkenlerin kümeleri ayırt etmede etkili olduğunu ifade etmektedir.

Number of Cases in each Cluster		
Cluster	1	701.000
	2	677.000
	3	717.000
	4	905.000
Valid		3000.000
Missing		.000

Şekil 78 Kümedeki Örnek Sayıları

Tablodaki verilere göre, kümeleme analizi sonucunda her bir kümeye ait gözlem sayıları şu şekilde dağılım göstermektedir:

- **1. Küme:** 701 gözlem
- **2. Küme:** 677 gözlem
- **3. Küme:** 717 gözlem
- **4. Küme:** 905 gözlem

Toplamda, 3000 geçerli gözlem analize dahil edilmiştir ve eksik veri bulunmamaktadır. En fazla gözlem sayısına sahip olan küme, 905 gözlem ile **4. Küme** olurken, en az gözlem sayısı ise 677 gözlem ile **2. Küme**de yer almaktadır. Bu dağılım, kümelerin homojen bir şekilde bölündüğünü ve belirgin bir dengesizlik olmadığını göstermektedir. Analiz edilen verilerde tüm

kümeler yeterli bir temsil gücüne sahiptir ve bu durum, kümeleme sonucunun genel geçerliliğini desteklemektedir.