



T.C.
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ
FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

LİSANS TEZİ

DIYABET HASTALIĞINA ETKİ EDEN FAKTÖRLERİN VERİ
ANALİZİ İLE İNCELENMESİ

Merve ZİREKOĞLU

Tez Danışmanı
Prof. Dr. Barış AŞIKGİL

İSTANBUL – 2024

ÖZ

Bu çalışma Kaggle'dan alınan DiabetesDataAnalysis adlı diyabet hastalarının bilgilerini içeren veri setinin, veri analizine uygun şekilde düzenlenip ilgili dönüşümlerin uygulanması ile hazırlanmıştır.

Çalışma; diyabetin tanımı, çeşitleri, risk faktörleri, belirtileri ve tanısı gibi temel konuları içeren bir giriş bölümü ile başlamaktadır. Ardından tanımlayıcı istatistikler, merkezi eğilim ölçüleri, değişim ölçüleri, grafik gösterimler, ilişki ölçüleri, normal dağılım testi, ki-kare bağımsızlık analizi, t-testi, varyans analizi, lojistik regresyon analizi gibi istatistiksel yöntemlere değinilmiştir. Son bölümde ise uygulama ile varsayım incelemesi yapılmıştır.

ABSTRACT

This study was prepared by organizing the data set containing the information of diabetes patients named DiabetesDataAnalysis obtained from Kaggle in accordance with the data analysis and applying the relevant transformations.

The study starts with an introductory section including basic topics such as definition, types, risk factors, symptoms and diagnosis of diabetes. Then, statistical methods such as descriptive statistics, measures of central tendency, measures of change, graphical representations, measures of association, normal distribution test, chi-square independence analysis, t-test, analysis of variance, logistic regression analysis are mentioned. In the last section, the application and assumption analysis were performed.

ÖNSÖZ

Çalışma boyunca bilgi ve tecrübelerini eksik etmeden her zaman yardımcı ve destek olan değerli tez danışmanım Mimar Sinan Güzel Sanatlar Üniversitesi İstatistik Bölüm Başkanı Prof. Dr. Barış AŞIKGİL'e sevgi, saygı ve teşekkürü borç bilirim.

İÇİNDEKİLER

1 GİRİŞ.....	1
1.1 Diyabet Nedir ?.....	1
1.2 Diyabet Çeşitleri	1
1.3 Diyabet Risk Faktörleri.....	1
1.4 Diyabet Belirtileri	1
1.5 Diyabet Tanısı	1
2 TANIMLAYICI İSTATİSTİKLER.....	2
2.1 Merkezi Eğilim Ölçüleri	2
2.1.1 Aritmetik Ortalama	2
2.1.2 Mod (Tepe Değer).....	2
2.1.3 Medyan (Ortanca).....	2
2.2 Değişim (Yayılım) Ölçüleri	2
2.2.1 Değişim Genişliği	2
2.2.2 Çeyrek Sapma	2
2.2.3 Varyans ve Standart Sapma.....	2
3 VERİ ÇÖZÜMLEMEDE GRAFİK GÖSTERİMLER	3
3.1 Kutu Grafiği (Boxplot)	3
3.2 Histogram Grafiği	3
3.3 Pasta (Daire) Grafiği.....	4
3.4 Q-Q Çizimleri	4
3.5 Saçılım Grafiği.....	5
4 İLİŞKİ ÖLÇÜLERİ.....	5
4.1 Pearson İlişki (Korelasyon) Katsayısı.....	5
5 NORMAL DAĞILIM.....	6
5.1 Normal Dağılım Testi	6
6 Kİ-KARE BAĞIMSIZLIK ANALİZİ.....	6
7 T-TESTİ.....	7
7.1 Tek Örneklem İçin T-Testi	7
7.2 Bağımsız İki Örneklem T-Testi.....	7
7.3 Bağımlı İki Örneklem T-Testi	8

8 VARYANS ANALİZİ.....	8
8.1 Tek Faktörlü Varyans Analizi (One-Way ANOVA)	8
9 LOJİSTİK REGRESYON ANALİZİ.....	9
9.1 Lojistik Regresyon Yöntemleri.....	9
9.1.1 İkili Lojistik Regresyon	9
9.1.2 Ordinal Lojistik Regresyon.....	10
9.1.3 Nominal Lojistik Regresyon.....	10
9.2 Lojistik Regresyon Avantajları ve Dezavantajları	10
10 UYGULAMA	11
10.1 Veri Seti Tanımı	11
10.1.1 Eksik Değer Kontrolü	12
10.1.2 Kategorik Değişkenleri Belirleme	13
10.2 Nitel Değişkenlerin Frekans Analizi	14
10.3 Nicel Değişkenlerin Tanımlayıcı İstatistikleri	17
10.4 Sürekli Değişkenlerin Normallik Varsayımı Analizi.....	19
10.5 Sürekli Değişkenlerin Histogram Grafikleri	21
10.6 Sürekli Değişkenlerin Kutu Grafikleri.....	22
10.7 Sürekli Değişkenlerin Q-Q Grafikleri.....	23
10.8 Sürekli Değişkenlerin Korelasyon Analizi.....	24
10.9 Ki-Kare Bağımsızlık Analizi.....	25
10.10 T Testi ve Varyans Analizi	29
10.11 Varyans Analizi	31
10.12 Lojistik Regresyon Analizi	39
10.13 Varsayım İncelemesi	42
10.14 Çoklu Doğrusal Bağlantı Sorununun İncelenmesi.....	43
10.15 Sağlam Kestirim Yöntemi ile Lojistik Regresyon Analizi.....	43
10.16 Model Denklemi	46
KAYNAKÇA.....	48

ŞEKİLLER TABLOSU

Şekil 3.1: Kutu Grafiği [8]	3
Şekil 3.2: Histogram Grafiği [9]	3
Şekil 3.3: Pasta Grafiği [10]	4
Şekil 3.4: Q-Q Çizimi [11]	4
Şekil 3.5: Saçılım Grafiği [12]	5
Şekil 4: Pearson İlişki Katsayıları ve Yorumları	5
Şekil 5: Çarpıklık ve Simetriklik [13]	6
Şekil 9: Sigmoid Eğrisi [19]	9
Şekil 10: Eksik Değer Kontrolü (R Programlama Çıktısı)	12
Şekil 10.2.1: İnsulin Değişkeni Oranları (R Programlama Çıktısı)	14
Şekil 10.2.2: İnsulin Değişkeni Pasta Grafiği (R Programlama Çıktısı)	14
Şekil 10.2.3: BMI Değişkeni Oranları (R Programlama Çıktısı)	15
Şekil 10.2.4: BMI Değişkeni Pasta Grafiği (R Programlama Çıktısı)	15
Şekil 10.2.5: Outcome Değişkeni Oranları (R Programlama Çıktısı)	16
Şekil 10.2.6: Outcome Değişkeni Pasta Grafiği (R Programlama Çıktısı)	16
Şekil 10.3.1: Pregnancies Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	17
Şekil 10.3.2: Glucose Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	17
Şekil 10.3.3: BloodPressure Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	17
Şekil 10.3.4: SkinThickness Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	18
Şekil 10.3.5: DiabetesPedigreeFunction Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	18
Şekil 10.3.6: Age Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)	19
Şekil 10.4: Normallik Testi (SPSS Çıktısı)	19
Şekil 10.5: Sürekli Değişkenlerin Histogram Grafikleri (R Programlama Çıktısı)	21
Şekil 10.6: Sürekli Değişkenlerin Kutu Grafikleri (R Programlama Çıktısı)	22
Şekil 10.7: Sürekli Değişkenlerin Q-Q Grafikleri (R Programlama Çıktısı)	23
Şekil 10.8: Sürekli Değişkenlerin Korelasyon Grafiği (R Programlama Çıktısı)	24
Şekil 10.9.1: İnsulin ve BMI Değişkenleri Çapraz Tablosu (SPSS Çıktısı)	25
Şekil 10.9.2: Kİ-Kare Testi (SPSS Çıktısı)	25
Şekil 10.9.3: İnsulin ve Outcome Değişkenleri Çapraz Tablosu (SPSS Çıktısı)	26
Şekil 10.9.4: Kİ-Kare Testi (SPSS Çıktısı)	27
Şekil 10.9.5: İnsulin ve Outcome Değişkenleri Çapraz Tablosu (SPSS Çıktısı)	27
Şekil 10.9.6: Kİ-Kare Testi (SPSS Çıktısı)	28
Şekil 10.10.1: SkinThickness – Outcome Değişkenleri Grup İstatistikleri (SPSS Çıktısı) ..	29
Şekil 10.10.2: SkinThickness – Outcome Bağımsız Örneklem Testi (SPSS Çıktısı)	29
Şekil 10.10.3: Glucose – Outcome Değişkenleri Grup İstatistikleri (SPSS Çıktısı)	30
Şekil 10.10.4: Glucose – Outcome Bağımsız Örneklem Testi (SPSS Çıktısı)	30
Şekil 10.11.1: BMI – DiabetesPedigreeFunction Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)	31
Şekil 10.11.2: Varyans Homojenliği Testi (SPSS Çıktısı)	32
Şekil 10.11.3: ANOVA (SPSS Çıktısı)	33
Şekil 10.11.4: Post Hoc Testi (SPSS Çıktısı)	33
Şekil 10.11.5: İnsulin – Glucose Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)	34
Şekil 10.11.6: Varyans Homojenliği Testi (SPSS Çıktısı)	34
Şekil 10.11.7: Robust Test (SPSS Çıktısı)	35
Şekil 10.11.8: Post Hoc Testi (SPSS Çıktısı)	36
Şekil 10.11.9: BMI – DiabetesPedigreeFunction Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)	36

Şekil 10.11.10: Varyans Homojenliği Testi (SPSS Çıktısı).....	37
Şekil 10.11.11: Robust Test (SPSS Çıktısı).....	37
Şekil 10.11.12: Post Hoc Testi (SPSS Çıktısı)	38
Şekil 10.12.1: Omnibus Testi (SPSS Çıktısı).....	39
Şekil 10.12.2: Bağımsız Değişkenlerin Modele Etkileri (SPSS Çıktısı).....	40
Şekil 10.12.3: Uyumluluk Testi (SPSS Çıktısı)	41
Şekil 10.13.1: Otokorelasyon İncelemesi (SPSS Çıktısı)	42
Şekil 10.13.2: Değişen Varyanslılık İncelemesi (SPSS Çıktısı).....	42
Şekil 10.14: VIF Değerleri (SPSS Çıktısı).....	43
Şekil 10.15.1: Omnibus Testi (SPSS Çıktısı).....	43
Şekil 10.15.2: Bağımsız Değişkenlerin Modele Etkileri (SPSS Çıktısı).....	44
Şekil 10.15.3: Uyumluluk Testi (SPSS Çıktısı)	45
Şekil 10.16: Parametre Tahminleri (SPSS Çıktısı).....	46

1 GİRİŞ

1.1 Diyabet Nedir ?

Yenilen besinler sindirildikten sonra vücutta bulunan enzimler sayesinde şekere yani glikoza parçalanıp kan akımı ile tüm vücuda yayılır. Şekerin, enerjiyi sağlayabilmesi için vücut hücrelerine girmesi gerekir. İnsülin ise pankreastan salgılanan ve şekerin kandan ayrılıp hücre içine girmesini sağlayan bir hormondur. Sağlıklı kişilerin her besin alımından sonra pankreasları, insülin üretir [1], [2].

Diyabet, halk arasında bilinen diğer ismi ile Şeker Hastalığı, pankreasın bazen yeteri kadar bazen hiç insülin üretememesinden veya hedef hücreler (kas, yağ ve karaciğer hücreleri) tarafından kullanılamamasından kaynaklanan; insülin salınımı, insülin etkisi gibi bozukluklar ile ortaya çıkan bir hastalıktır [2].

1.2 Diyabet Çeşitleri

Vücutta insülin salgısının hiç olmaması veya yok sayılacak kadar az olması durumu, Tip 1 Diyabet; vücutta insülin salgılanmasına rağmen miktarının ya da etkisinin yetersiz olması durumu ise Tip 2 Diyabet olarak adlandırılır [2].

1.3 Diyabet Risk Faktörleri

Diyabetin olası risk faktörleri; obezite, sağlıksız beslenme, aile öyküsü, yüksek tansiyon, hareketsiz hayat olarak tanımlanmıştır [1], [3].

1.4 Diyabet Belirtileri

Tip 1 Diyabet'in; bulantı, kusma, kilo kaybı, karın ağrısı gibi semptomları vardır. Tip 2 Diyabet'in ise; yorgunluk, yaraların geç iyileşmesi, ağız kuruluğu, çok su içmek, sık ve bol miktarda idrar yapmak, kilo almak veya zayıflamak gibi semptomları vardır [3], [4].

1.5 Diyabet Tanısı

Sağlıklı kişilerde açlık kan şekeri düzeyi, ortalama 70-100 mg/Dl arasındadır. Bu ortalamanın 126 mg/Dl üzerinde olması, tanı koymak için yeterlidir. Ancak 70-126 mg/Dl arasında ise tokluk kan şekeri de bakılır ve 200 mg/Dl üzerinde olması Diyabetin, 140-199 mg/Dl aralığında olması ise Gizli Şekerin göstergesi olur [5].

2 TANIMLAYICI İSTATİSTİKLER

2.1 Merkezi Eğilim Ölçüleri

2.1.1 Aritmetik Ortalama

Bir serideki terimlerin toplamlarının, terim sayısına bölünmesi ile elde edilir [6].

2.1.2 Mod (Tepe Değer)

Veri setinde en çok tekrarlanan yani en yüksek frekansa sahip olan değerdir [6].

2.1.3 Medyan (Ortanca)

Veri grubundaki değerler küçükten büyüğe sıralandığında tam ortada kalan değere denir. Kitledeki birim sayısı çok ise veriyi özetlemek için medyan kullanılabilir [6]

2.2 Değişim (Yayılım) Ölçüleri

Verilerdeki değişkenliğin açıklanması, dağılımın incelenmesi, homojenliğin incelenip karşılaştırılması gibi incelemelerde kullanılır [7].

2.2.1 Değişim Genişliği

Veri setindeki en büyük değer ile en küçük değer arasındaki farktır. R ile gösterilir. Uç değerlere duyarlıdır [7].

2.2.2 Çeyrek Sapma

Veri setinde ortalama yerine ortancanın kullanılması gerektiği durumlarda, veri setinde uç değerlerin bulunması durumunda değişim genişliği yerine çeyrek sapma kullanılır.

Q: Çeyrek Sapma, Q1: 1. çeyreklik, Q3: 2. çeyreklik olmak üzere:

$Q = (Q3 - Q1) / 2$ formülü ile hesaplanır.

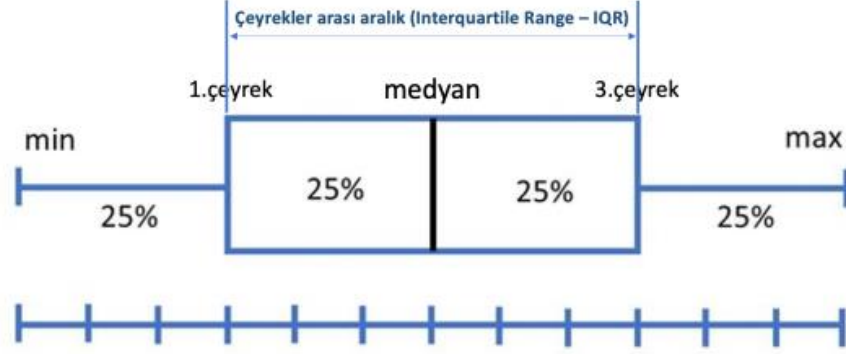
2.2.3 Varyans ve Standart Sapma

Varyans, veri setindeki her bir değer ile ortalamadan ayrılışlarının kareler toplamının, gözlem sayısının bir eksiğine bölünmesi ile bulunur. Standart sapma ise varyansın kareköküdür.

3 VERİ ÇÖZÜMLEMEDE GRAFİK GÖSTERİMLER

3.1 Kutu Grafiği (Boxplot)

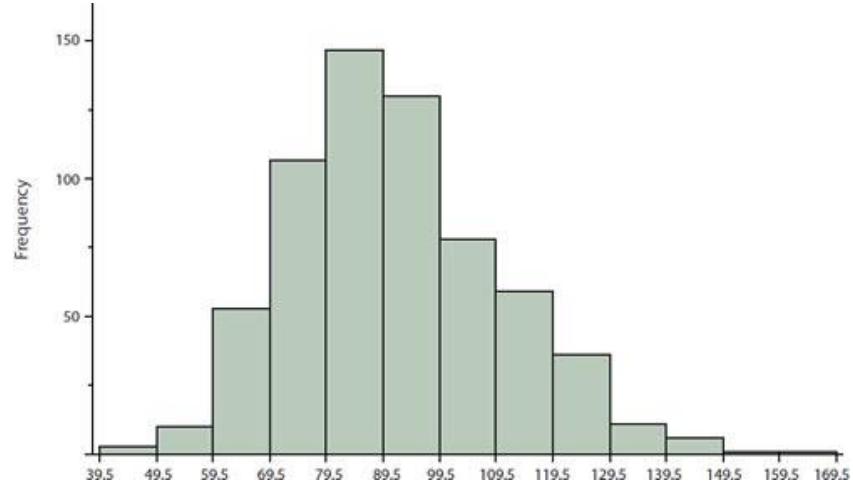
Kutu grafikleri verinin dağılımı, konumu, çarpıklığı, uç değerleri hakkında bilgi verir. 5 nokta özet değerleri ile oluşturulur.



Şekil 3.1: Kutu Grafiği [8]

3.2 Histogram Grafiği

Histogramlar verinin konumu, yayılımı, çarpıklığı, anormal davranışları hakkında bilgi verir. Sürekli değişkenlerin frekans dağılımlarında kullanılır.



Şekil 3.2: Histogram Grafiği [9]

3.3 Pasta (Daire) Grafiđi

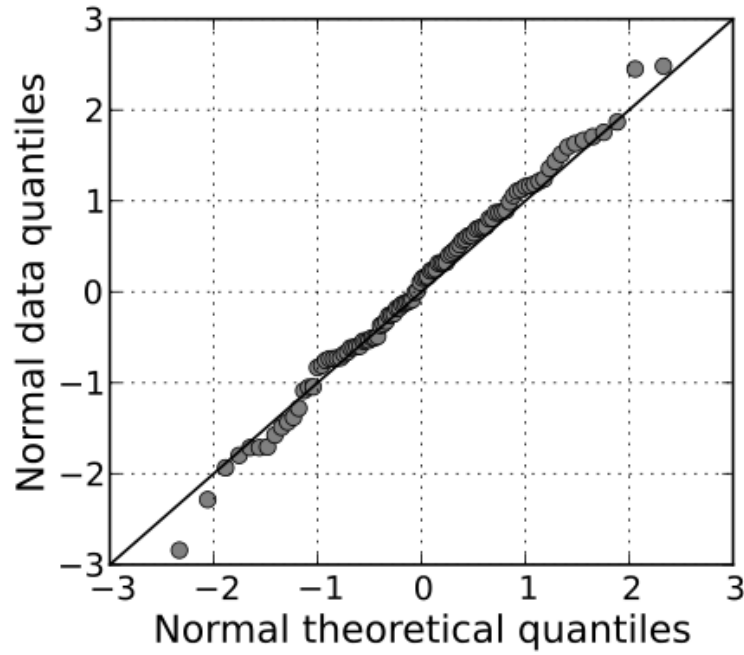
Veri setindeki gözlem değęerlerinin frekanslarını, 360 dereceye tamamlanan dilimler ile gösterir.



Şekil 3.3: Pasta Grafiđi [10]

3.4 Q-Q Çizimleri

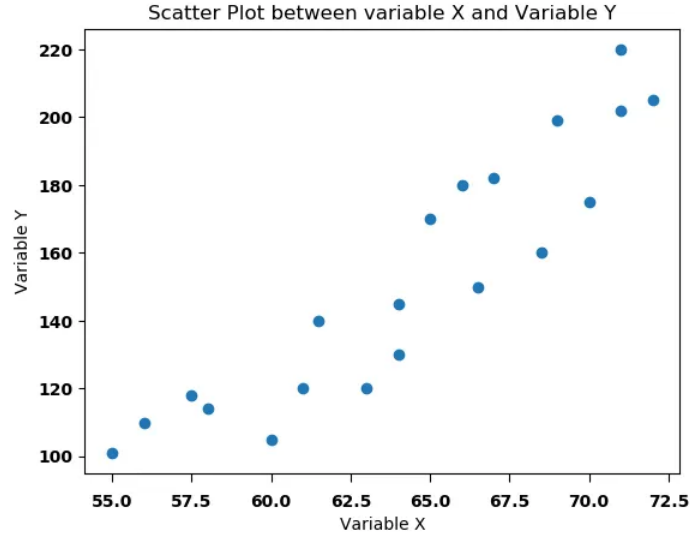
Bir veri setinin dağılımını kuramsal bir dağılımla karşılaştırmak, iki rastgele örneklemin aynı dağılımdan gelip gelmediđini araştırmak için kullanılır. Küçük veri setleri için uygun olmayabilir [7].



Şekil 3.4: Q-Q Çizimi [11]

3.5 Saçılım Grafiği

İki nicel değişken arasındaki ilişki, homojenlik, aykırı ve uç değer varlığı hakkında bilgi verir.



Şekil 3.5: Saçılım Grafiği [12]

4 İLİŞKİ ÖLÇÜLERİ

4.1 Pearson İlişki (Korelasyon) Katsayısı

Nicel bir değişkende olan değişimin, diğer değişkenlerdeki etkisi hakkında bilgi verir. Pearson İlişki Katsayısı -1 ile +1 arasında yer alır.

Pearson İlişki Katsayısı	İlişki Yorumu
< 0.25	Çok zayıf ilişki ya da ilişki yok
0.26 - 0.49	Zayıf ilişki
0.50 - 0.69	Orta düzeyde ilişki
0.70 - 0.89	Yüksek ilişki
0.90 - 1.00	Çok Yüksek İlişki

Şekil 4: Pearson İlişki Katsayıları ve Yorumları

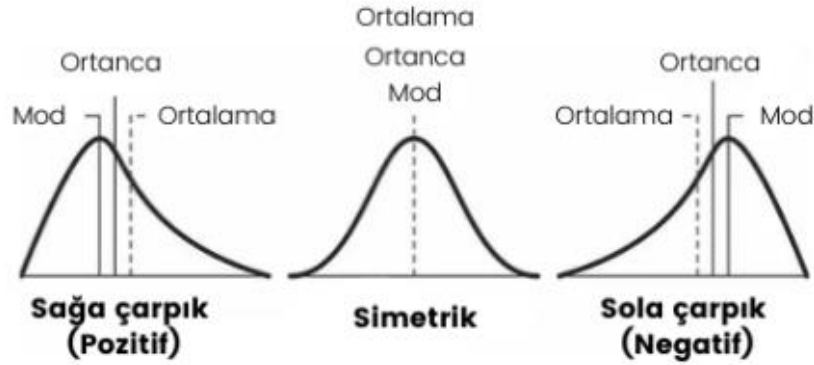
5 NORMAL DAĞILIM

Gauss Dağılımı veya Gauss Eğrisi olarak da bilinen Normal Dağılım, ortalaması 0 ve varyansı 1 olan simetrik yapıya sahip bir dağılımdır.

Simetrik yapıda: Ortalama = Mod = Medyan şeklindedir.

Sağa çarpık dağılımda: Ortalama > Medyan > Mod

Sola çarpık dağılımda: Mod > Medyan > Ortalama şeklindedir.



Şekil 5: Çarpıklık ve Simetriklik [13]

5.1 Normal Dağılım Testi

Normal dağılım için en çok tercih edilen testler arasında Kolmogorov-Smirnov ve Shapiro-Wilk testleri gelmektedir. Hipotezler ise aşağıdaki şekildedir:

H_0 : Veri seti normal dağılıma uygundur.

H_a : Veri seti normal dağılıma uygun değildir.

Hipotezlerin ardından test için p değeri incelenir. p değeri, anlamlılık değerinden küçük ise H_0 hipotezi reddedilir. p değeri, anlamlılık değerinden büyük ise H_0 hipotezi reddedilemez. Yani veriler normal dağılıma uygundur. H_0 hipotezinin kabul edilmediği durumlarda, normalliği yakalamak için çeşitli dönüşümler uygulanabilir.

6 Kİ-KARE BAĞIMSIZLIK ANALİZİ

İki değişken arasında ilişki, bağımlılık olup olmadığını test eder. Değişkenlerin tipleri nitel-nitel, nitel-nicel, nicel-nicel olabilir. Gruplar birbirinden bağımsız olmalıdır [14]. Hipotezler ise aşağıdaki şekildedir:

H_0 : İki değişken birbirinden bağımsızdır.

H_a : İki değişken birbirine bağımlıdır.

Formülü:

$$\chi^2 = \sum_{i=1}^k \frac{(G_i - B_i)^2}{B_i} \text{ şeklindedir.}$$

G_i : Gözlenen Frekanslar

B_i : Beklenen Frekanslar

k : Toplam Gözlem Sayısı

χ^2 hesaplandıktan sonra karar aşamasına geçilir. Örneklem sayısı 30'dan büyükse Z tablosundan, örneklem sayısı 30'dan küçükse χ^2 tablosundan yararlanılarak karar verilir. $\chi^2 < \chi^2_{\alpha}$ ise H_0 hipotezi reddedilemez. Ters durumda ise H_0 hipotezi reddedilir.

7 T-TESTİ

İki örneklem grubu arasında ortalama farkı olup olmadığını test eder [15]. 3 farklı t testi vardır.

7.1 Tek Örneklem İçin T-Testi

Örneklem ortalaması ile varsayımsal bir değer arasında istatistiksel olarak anlamlı bir fark olup olmadığını test etmek için kullanılır [16].

Anakütle standart sapması biliniyorsa Z istatistiği,

Anakütle standart sapması bilinmiyorsa ve $n > 30$ ise Z istatistiği,

Anakütle standart sapması bilinmiyor ve $n < 30$ ise T istatistiği, kullanılır. Hipotezler ise aşağıdaki şekildedir:

H_0 : Örneklem ortalaması ile tahmin edilen ortalama arasında fark yoktur.

H_a : Örneklem ortalaması ile tahmin edilen ortalama arasında fark vardır.

Hesaplanan değer < tablo değeri ise H_0 hipotezi reddedilemez. Ters durumda ise H_0 hipotezi reddedilir.

7.2 Bağımsız İki Örneklem T-Testi

İki grup ortalaması arasında karşılaştırma yapılmak istenildiğinde kullanılır. Testin uygulanabilmesi için normallik ve varyans homojenliği varsayımları sağlanmalıdır [16]. Hipotezler ise aşağıdaki şekildedir:

H_0 : İki örneklemin ortalamaları birbirine eşittir.

H_a : İki örneklemin ortalamaları birbirine eşit değildir.

Hesaplanan deęer < tablo deęeri ise H_0 hipotezi reddedilemez. Tersı durumda ise H_0 hipotezi reddedilir.

7.3 Baęımlı İki Örnekleme T-Testi

Baęımlı iki grup ortalaması arasında karřılařtırma yapılmak istenildięinde kullanılır. Aynı kitleye iki farklı uygulama yapıldıęında ve sonuçları incelendięinde buna baęımlı gruplar denir [16]. Hipotezler ise ařaęıdaki řekildedir:

H_0 : Olay, örnekleme üzerinde etkili deęildir.

H_a : Olay, örnekleme üzerinde etkilidir.

Hesaplanan deęer < tablo deęeri ise H_0 hipotezi reddedilemez. Tersı durumda ise H_0 hipotezi reddedilir.

8 VARYANS ANALİZİ

8.1 Tek Faktörlü Varyans Analizi (One-Way ANOVA)

Bir veya daha fazla baęımsız deęiřkenin bir baęımlı deęiřken ortalamaları arasında istatistiksel olarak anlamlı fark olup olmadıęını test eder [17]. Levene veya Bartlett testi ile uygulanabilir. Uygulanabilmesi için normallik, homojen varyanslılık, baęımsızlık gibi varsayımların saęlanması gerekmektedir. Hipotezler ise ařaęıdaki řekildedir:

H_0 : Gruplar arası varyans eřittir.

H_a : Grup arası varyans eřit deęildir.

Hesaplanan deęer < tablo deęeri ise H_0 hipotezi reddedilemez. Tersı durumda ise H_0 hipotezi reddedilir.

Ardından grup ortalamaları farkı için hipotezler test edilir.

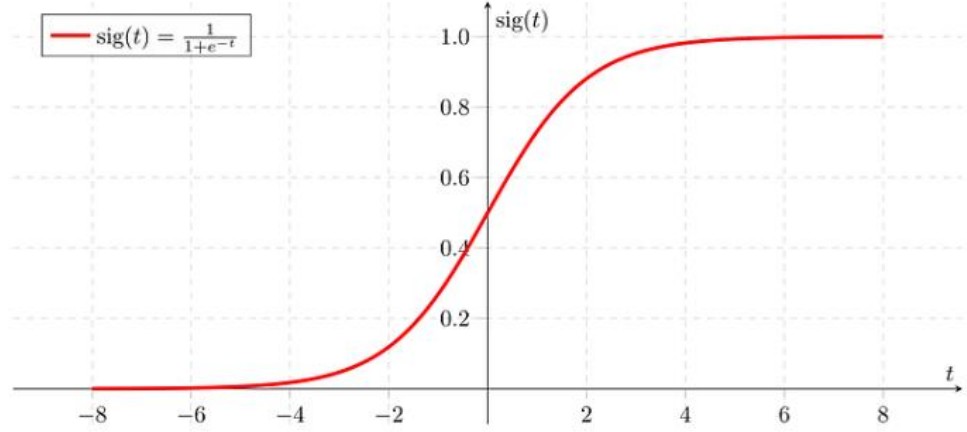
H_0 : Grupların ortalamaları arasında anlamlı fark yoktur.

H_a : Grupların ortalamaları arasında anlamlı fark vardır.

Hesaplanan deęer < tablo deęeri ise H_0 hipotezi reddedilemez. Tersı durumda ise H_0 hipotezi reddedilir.

9 LOJİSTİK REGRESYON ANALİZİ

Lojistik Regresyon, bağımlı değişkenin kategorik değişken olduğu bir sınıflandırma türüdür. Optimum değeri Maksimum Olabilirlik ile çizer [18]. Sınıflandırma yapmak için Sigmoid (Lojistik) Fonksiyonu kullanır ve bu fonksiyon “S” şeklinde bir eğridir. Sigmoid fonksiyonu, verileri 0 ve 1 arasına sıkıştırmak için kullanılan bir fonksiyondur [19].



Şekil 9: Sigmoid Eğrisi [19]

p: Olayın gözlenme olasılığı olmak üzere Lojistik Regresyon Modeli aşağıdaki gibi kurulur:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Lojistik Regresyon Analizinde 3 farklı yöntem vardır.

9.1 Lojistik Regresyon Yöntemleri

9.1.1 İkili Lojistik Regresyon

Bağımlı değişkeni “binary” türünde yani, 2 kategoriden oluşan regresyon modellerine denir. İkili değişken “0” ve “1” gibi sayılarla kodlanabilir. Örneğin diyabet hastaları üzerinde yapılan bir çalışmada 0, “Diyabeti Yok”; 1, “Diyabeti Var” anlamına

gelmektedir. Aynı şekilde sigara kullanımı üzerine olan bir çalışmada 0, “Sigara kullanmıyor” anlamına gelmekte iken 1, “Sigara kullanıyor” anlamına gelmektedir.

9.1.2 Ordinal Lojistik Regresyon

Bağımlı değişken en az 3 kategoriden oluşuyor olup kategoriler, doğal sınıflama yapısına uymalıdır.

Örneğin Vücut Kitle İndeksine(VKİ) yönelik bir araştırmada VKİ;

- 18, 5 kg/m² ‘nin altında olan kişiler “İdeal kilonun altında”,
- 18, 5 kg/m² ile 24, 9 kg/m² arasında olan kişiler “İdeal kiloda”,
- 25 kg/m² ile 29, 9 kg/m² arasında olan kişiler “İdeal kilonun üstünde”,
- 30 kg/m² ile 39, 9 kg/m² arasında olan kişiler “İdeal kilonun çok üstünde (Obez)”,
- 40 kg/m² üzerinde olan kişiler “İdeal kilonun çok üstünde (Morbid Obez)”,

olarak adlandırılır. Bu sıralamaya göre sırasıyla “Zayıf” = 0, “İdeal” = 1, “Kilolu” = 2, “Obez” = 3, “Morbid Obez” = 4 şeklinde kategorizasyon sağlanmalıdır.

9.1.3 Nominal Lojistik Regresyon

Bağımlı değişken en az 3 kategoriden oluşuyor olup kategoriler, isimsel ölçekli olmalıdır. Örneğin kitap türleri veya ders isimlerinin kategorize edilmesi Nominal Lojistik Regresyon kapsamında olmaktadır.

9.2 Lojistik Regresyon Avantajları ve Dezavantajları

Lojistik regresyonun uygulanması ve yorumlanması kolaydır. Veri seti doğrusal olarak ayrılabiliriyorsa oldukça iyi performans gösterir. Aşırı Öğrenme ’ye (Overfitting) daha az meyillidir ama büyük veri setlerinde Aşırı Öğrenme olabilir [18].

10 UYGULAMA

10.1 Veri Seti Tanımı

Bu çalışma Kaggle'dan alınan DiabetesDataAnalysis [20] adlı veri setinin veri analizine uygun şekilde düzenlenip ilgili dönüşümlerin uygulanması ile hazırlanmıştır. Düzenlenen veri seti, hastaların 9 klinik özelliğini içermektedir. Çalışmanın amacı ise 8 özellik (değişken) yardımıyla, kişilerin Diyabet hastası olup olmadığını saptamaktır.

Veri seti 9 değişken ve 670 gözlemden oluşmaktadır.

Değişkenler;

- **Pregnancies:** Kişilerin, veri toplanmadan önceki doğum yapma sayıları.
- **Glucose:** Şeker seviyesi.
- **BloodPressure:** Kan basıncı (mm/Hg).
- **SkinThickness:** Cilt kalınlığı (mm).
- **Insulin:** İnsülin değeri (2 saatlik serum insülini).
- **BMI:** Vücut kitle indeksi ($\text{kg}/(\text{boy})^2$).
- **DiabetesPedigreeFunction:** Diyabet soyağacı işlevi (Aile geçmişine dayalı olarak diyabet olasılığını puanlayan bir işlev).
- **Age:** Yaş.
- **Outcome:** Sonuç (0: Diyabeti yok, 1: Diyabeti var.).

Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age değişkenleri sayısal değişkenlerdir. Outcome (çıktı) değişkeni ise 0 ve 1'lerden oluşup kişilerin diyabeti olup olmadığını gösteren bağımlı değişkendir.

Veri analizi uygulamaları, R Programlama (R version 4.2.2) ve IBM SPSS 26 kullanılarak hazırlanmıştır.

10.1.1 Eksik Değer Kontrolü

Pregnancies	Glucose	BloodPressure	SkinThickness
0	0	0	0
Insulin	BMI	DiabetesPedigreeFunction	Age
0	0	0	0
Outcome			
0			

Şekil 10: Eksik Değer Kontrolü (R Programlama Çıktısı)

Şekil 10’da görüldüğü üzere veri seti, eksik değer içermemektedir. Ancak sütun değerleri tek tek incelendiğinde “0” değerleri tespit edilmiştir.

Eksik değerlerin doldurulmasında Regresyon Yöntemi kullanılmıştır. Bu yöntemde eksik değerlerin doldurulacağı değişken bağımlı, diğer değişken veya değişkenler ise bağımsız olarak ele alınmaktadır. Sürekli değişkenleri için En Küçük Kareler Yöntemi kullanılır [7].

- Pregnancies değişkeninde yer alan “0” değerleri, eksik değer değildir. Çünkü geçmiş hamilelik durumunu belirten Pregnancies değişkeninin “0” değerini alması, ilgili gözlemin hiç doğum yapmadığı bilgisini verir. Ayrıca, ilgili gözlemin cinsiyetinin “Erkek” olma ihtimalini de beraberinde getirir.
- Glucose değişkeninde yer alan “0” değerleri incelenmemiştir. Çünkü veri seti Diyabet hastalarına ait olduğu için Glucose değerleri, normal bireylerden farklı sonuç verebilir.
- BloodPressure değişkeninde yer alan “0” değerleri Regresyon Yöntemi kullanılarak diğer değişkenlerin yardımı ile doldurulmuştur.
- SkinThickness değişkeninde yer alan “0” değerleri Regresyon Yöntemi kullanılarak diğer değişkenlerin yardımı ile doldurulmuştur.
- Insulin değişkeninde yer alan “0” değerleri incelenmemiştir. Çünkü veri seti Diyabet hastalarına ait olduğu için Insulin değerleri, normal bireylerden farklı sonuç verebilir.
- BMI değişkeninde yer alan “0” değerleri Regresyon Yöntemi kullanılarak diğer değişkenlerin yardımı ile doldurulmuştur.
- DiabetesPedigreeFunction değişkeninde “0” değeri bulunmamaktadır.
- Age değişkeninde “0” değeri bulunmamaktadır.
- Outcome değişkeni zaten 0 ve 1’lerden oluşup “0” değerleri, diyabetin olmadığı anlamına gelmektedir.

10.1.2 Kategorik Değişkenleri Belirleme

Veri setinde yer alan bağımsız değişkenlerin tamamı nicel değişken türündedir. Veri setini daha kapsamlı incelemek amacıyla İnsulin ve BMI değişkenleri, kategorik değişkene çevrilecektir. Ayrıca çıktı değişkeni olan Outcome değişkeni de kategorik değişken olarak tanımlanacaktır.

- İnsulin değişkeninin kategorik değişkene çevrilmesi.

İnsülin değişkeni, tokluk kan şekerini temsil etmekteydi. Tokluk kan şekerinin;

- 140 mg/Dl ‘den az olması “Normal”,
- 140-199 mg/Dl aralığında olması “Gizli Şeker”,
- 200 mg/Dl üzerinde olması “Diyabet”, göstergesidir.

Bu sınırlara göre İnsülin değerleri kategorize edilecektir.

- BMI değişkeninin kategorik değişkene çevrilmesi.

Uzmanlara göre [21] BMI değeri;

- 18, 5 kg/m² ‘nin altında olan kişiler “İdeal kilonun altında”,
- 18, 5 kg/m² ile 24, 9 kg/m² arasında olan kişiler “İdeal kiloda”,
- 25 kg/m² ile 29, 9 kg/m² arasında olan kişiler “İdeal kilonun üstünde”,
- 30 kg/m² ile 39, 9 kg/m² arasında olan kişiler “İdeal kilonun çok üstünde (obez)”,
- 40 kg/m² üzerinde olan kişiler “İdeal kilonun çok üstünde (morbid obez)”, olarak adlandırılır.

Bu bilgilere göre BMI değişkeni, sırasıyla “Zayıf”, “İdeal”, “Kilolu”, “Obez”, “Morbid Obez” olarak kategorize edilecektir.

- Outcome değişkeninin kategorik değişkene çevrilmesi.

Bağımlı değişken olan Outcome değişkeninde yer alan 0 değeri, hastanın diyabetinin olmadığı; 1 değeri ise hastanın diyabeti olduğu anlamına gelmektedir.

Bu bilgiye göre Outcome değişkeni; 0 için “Diyabeti yok”, 1 için “Diyabeti var” şeklinde kategorize edilecektir.

10.2 Nitel Değişkenlerin Frekans Analizi

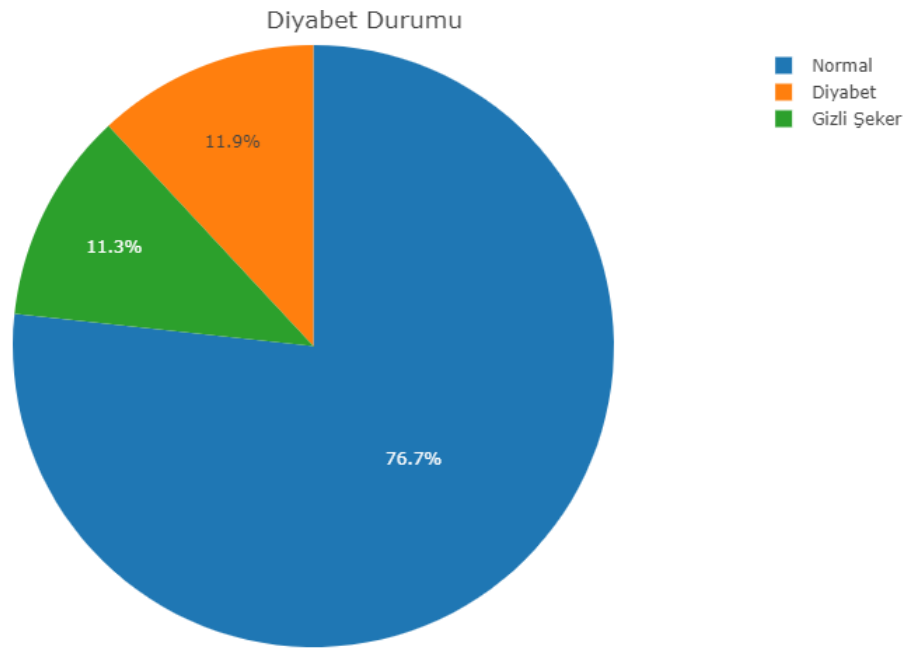
Bu başlık altında kategorik değişkenlerin frekansları, oranları ve pasta grafikleri incelenecektir.

- İnsulin Değişkeni

Kategori <chr>	Frekanslar <dbl>	Oranlar <dbl>
Diyabet	80	0.1194030
Gizli Şeker	76	0.1134328
Normal	514	0.7671642

Şekil 10.2.1: İnsulin Değişkeni Oranları (R Programlama Çıktısı)

Tablo incelendiğinde gözlemlerin tokluk kan şekeri değerlerine göre; 80 kişinin diyabet, 76 kişinin gizli şeker ve 514 kişinin normal olarak adlandırılan sınırlarda yer aldığı görülmektedir.



Şekil 10.2.2: İnsulin Değişkeni Pasta Grafiği (R Programlama Çıktısı)

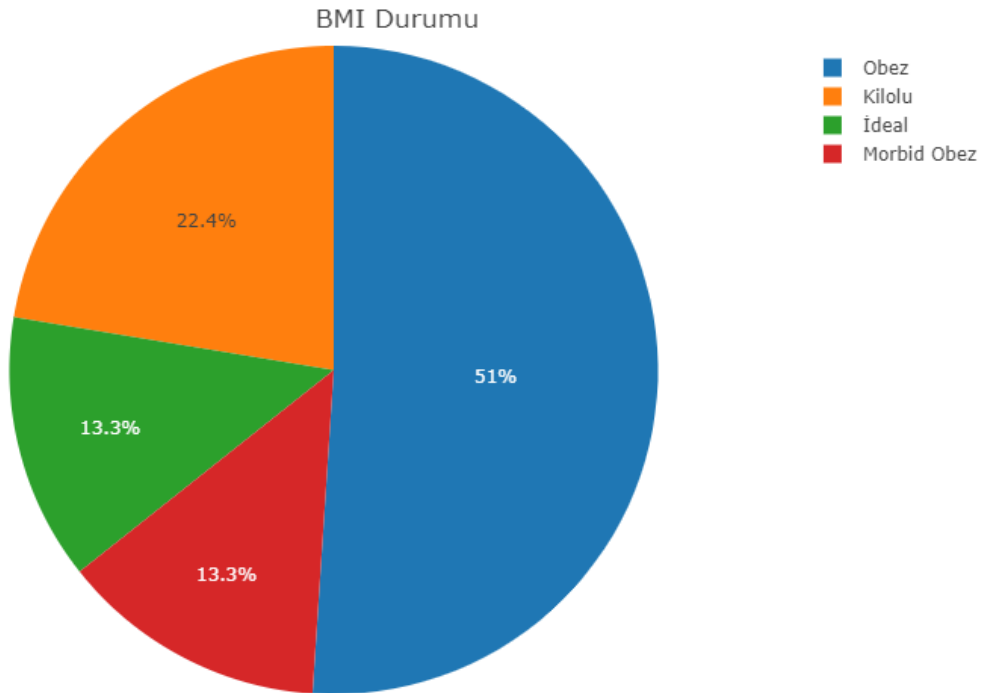
Bu sonuçlara göre kişilerin %11.9'u diyabet, %11.3'ü gizli şeker, %76.7'sinin normal kategorilerinde yer almaktadır.

- BMI Değişkeni

Kategori <chr>	Frekanslar <dbl>	Oranlar <dbl>
İdeal	89	0.1328358
Kilolu	150	0.2238806
Morbid Obez	89	0.1328358
Obez	342	0.5104478

Şekil 10.2.3: BMI Değişkeni Oranları (R Programlama Çıktısı)

Tablo incelendiğinde 88 kişinin ideal kiloda, 149 kişinin kilolu, 88 kişinin morbid obez ve 344 kişinin obez olduğu görülmektedir.



Şekil 10.2.4: BMI Değişkeni Pasta Grafiği (R Programlama Çıktısı)

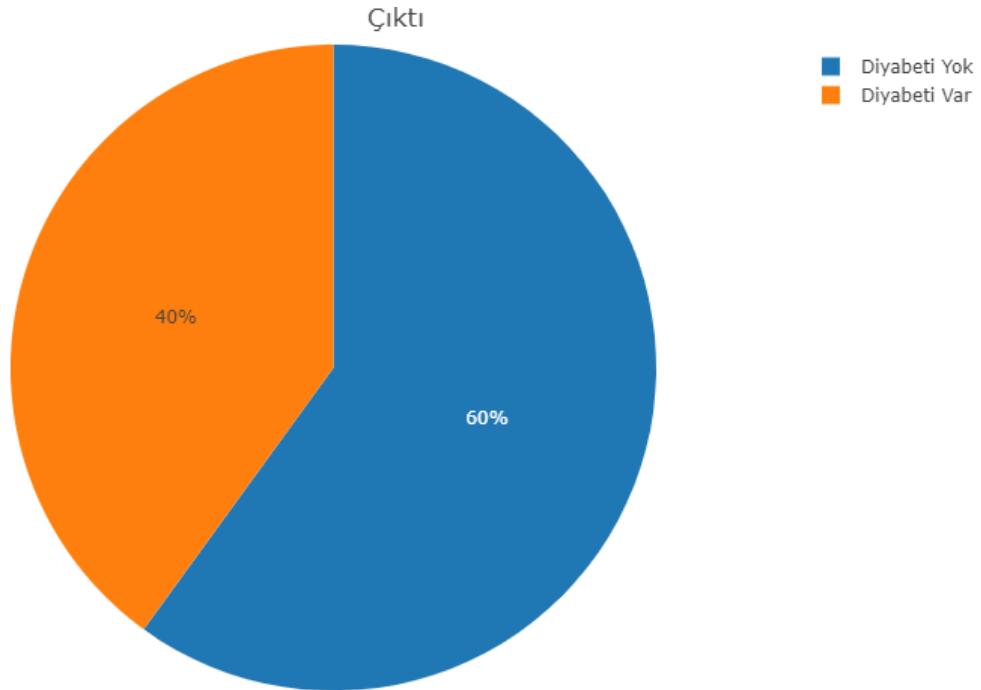
Bu sonuçlara göre kişilerin %13.3'ünün ideal kiloda, %22.4'ünün kilolu, %13.3'ünün morbid obez ve %51'inin obez kategorilerinde yer aldığı görülmektedir.

- Outcome Değişkeni

Kategori <chr>	Frekanslar <dbl>	Oranlar <dbl>
Diyabeti Var	268	0.4
Diyabeti Yok	402	0.6

Şekil 10.2.5: Outcome Değişkeni Oranları (R Programlama Çıktısı)

Tablo incelendiğinde 268 kişinin diyabet hastası olduğunu, 401 kişinin ise diyabet hastası olmadığı görülmektedir.



Şekil 10.2.6: Outcome Değişkeni Pasta Grafiği (R Programlama Çıktısı)

Bu sonuçlara göre kişilerin %40.1'i diyabeti olan, %59.9'u ise diyabeti olmayan kategorilerinde yer almaktadır.

10.3 Nicel Değişkenlerin Tanımlayıcı İstatistikleri

- Pregnancies Değişkeni

Descriptive Statistics												
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Pregnancies	670	17	0	17	3,89	,131	3,390	11,493	,873	,094	,101	,189
Valid N (listwise)	670											

Şekil 10.3.1: Pregnancies Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Pregnancies değişkeninin minimum değerinin 0, maksimum değerinin 17 olduğu görülmektedir. Ortalaması 3.89 ± 0.131 olarak bulunmuştur.

- Glucose Değişkeni

Descriptive Statistics												
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Glucose	670	199	0	199	122,42	1,230	31,827	1012,938	,249	,094	,387	,189
Valid N (listwise)	670											

Şekil 10.3.2: Glucose Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Glucose değişkeninin minimum değerinin 0, maksimum değerinin 199 olduğu görülmektedir. Ortalaması 122.42 ± 1.230 olarak bulunmuştur. Çarpıklık katsayısı 0.249 ve basıklık katsayısı 0.387’dir. Bu bilgilerle Glucose değişkeni dağılımının basık ve hafif sağa çarpık olduğu yorumu yapılabilir.

- BloodPressure Değişkeni

Descriptive Statistics												
	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
BloodPressure	670	92	30	122	72,48	,476	12,329	151,999	,237	,094	,821	,189
Valid N (listwise)	670											

Şekil 10.3.3: BloodPressure Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

BloodPressure değişkeninin minimum değerinin 30, maksimum değerinin 122 olduğu görülmektedir. Ortalaması 72.48 ± 0.476 olarak bulunmuştur. Çarpıklık katsayısı 0.237 ve basıklık katsayısı 0.821'dir. Bu bilgilerle BloodPressure değişkeni dağılımının basık ve hafif sağa çarpık olduğu yorumu yapılabilir.

- SkinThickness Değişkeni

Descriptive Statistics												
	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Error Std. Error	Std. Deviation Statistic	Variance Statistic	Skewness Statistic	Std. Error Std. Error	Kurtosis Statistic	Std. Error Std. Error
SkinThickness	670	92	7	99	29,12	,397	10,267	105,418	,649	,094	2,555	,189
Valid N (listwise)	670											

Şekil 10.3.4: SkinThickness Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

SkinThickness değişkeninin minimum değerinin 7, maksimum değerinin 99 olduğu görülmektedir. Ortalaması 29.12 ± 0.397 olarak bulunmuştur. Çarpıklık katsayısı 0.649 ve basıklık katsayısı 2.555'tir. Bu bilgilerle SkinThickness değişkeni dağılımının sivriye yakın ve sağa çarpık olduğu yorumu yapılabilir.

- DiabetesPedigreeFunction Değişkeni

Descriptive Statistics												
	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Error Std. Error	Std. Deviation Statistic	Variance Statistic	Skewness Statistic	Std. Error Std. Error	Kurtosis Statistic	Std. Error Std. Error
DiabetesPedigreeFunction	670	2,342	,078	2,420	,47387	,012618	,326601	,107	1,892	,094	5,847	,189
Valid N (listwise)	670											

Şekil 10.3.5: DiabetesPedigreeFunction Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

DiabetesPedigreeFunction değişkeninin minimum değerinin 0.078, maksimum değerinin 2.420 olduğu görülmektedir. Ortalaması 0.47387 ± 0.012618 olarak bulunmuştur. Çarpıklık katsayısı 1.892 ve basıklık katsayısı 5.847'dir. Bu bilgilerle DiabetesPedigreeFunction değişkeni dağılımının sivri olduğu yorumu yapılabilir.

- Age Değişkeni

Descriptive Statistics												
	N Statistic	Range Statistic	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Error Std. Error	Std. Deviation Statistic	Variance Statistic	Skewness Statistic	Std. Error Std. Error	Kurtosis Statistic	Std. Error Std. Error
Age	670	51	21	72	33,52	,452	11,710	137,123	1,059	,094	,373	,189
Valid N (listwise)	670											

Şekil 10.3.6: Age Değişkeni Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Age değişkeninin minimum değerinin 21, maksimum değerinin 72 olduğu görülmektedir. Ortalaması 33.52 ± 0.452 olarak bulunmuştur. Çarpıklık katsayısı 1.059 ve basıklık katsayısı 0.373'tür. Bu bilgilerle Age değişkeni dağılımının basık ve sağa çarpık olduğu yorumu yapılabilir.

10.4 Sürekli Değişkenlerin Normallik Varsayımı Analizi

Veri setinde incelenen değişkenlerin, normallik varsayımını sağlamadıkları için dönüşüm uygulamasına karar verilmiştir. Glucose , DiabetesPedigreeFunction ve Age değişkenlerinin histogramları göz önünde bulundurularak Logaritmik dönüşüm uygun bulunmuştur.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
pregnancies_kok	,104	667	,000	,952	667	,000
glucose_log	,030	667	,200*	,990	667	,000
bloodPressure_kok	,054	667	,000	,991	667	,000
skinThickness_kok	,053	667	,000	,990	667	,000
diabetesPedigreeFunction_log	,054	667	,000	,993	667	,003
age_log	,115	667	,000	,932	667	,000

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

a. Lilliefors Significance Correction

Şekil 10.4: Normallik Testi (SPSS Çıktısı)

$\alpha = 0.05$ düzeyinde, veri setinde yer alan sürekli değişkenlerin normal dağılım ile dağılıp dağılmadıkları test edilecektir.

Gözlem sayısı fazla olduğu dönüşümlerden sonra yaklaşık normal kabul edilerek analizlere devam edilecektir.

Hipotezler:

H_0 : Veri seti normal dağılmaktadır.

H_a : Veri seti normal dağılmamaktadır.

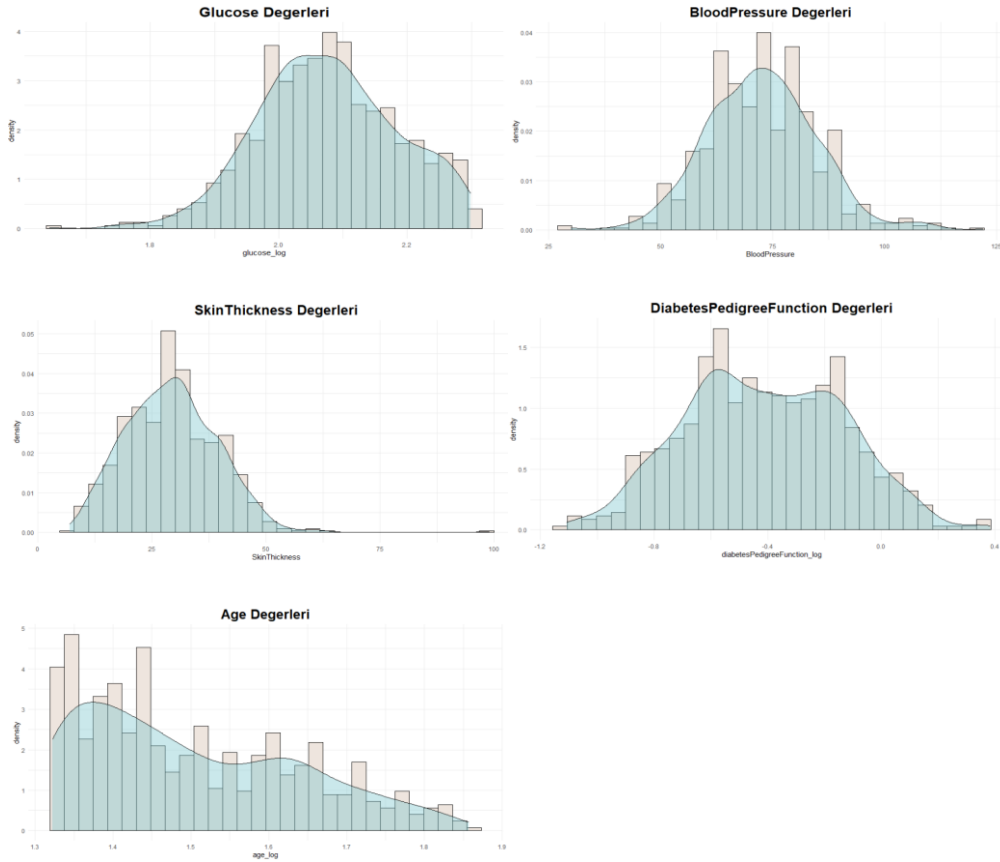
Karar: Veri setinin değişken sayısı, 50'den büyük olduğu için Kolmogorov-Smirnov testi ile karar verilecektir.

Yorum: Değişkenlerin significance değerlerine bakıldığında yalnızca Glucose değişkeninin, $\alpha = 0.05$ değerinden büyük olduğu görülmektedir. Bu nedenle H_0 , yani “Veri seti normal dağılmaktadır.” hipotezi Glucose değişkeni için %95 güven düzeyi ile reddedilemez.

Diğer değişkenlerin significance değerlerine bakıldığında, $\alpha = 0.05$ değerinden küçük olduğu görülmektedir. Bu nedenle H_0 , yani “Veri seti normal dağılmaktadır.” hipotezi diğer değişkenler için %95 güven düzeyi ile reddedilir.

Diğer değişkenler için farklı dönüşümler de denenmiş olup sonuç alınamamıştır. Bu sebeple tüm değişkenlerin normal dağıldığı varsayımı altında analize devam edilecektir. DiabetesPedigreeFunction ve Age değişkenlerinin histogramları incelendiğinde logaritmik dönüşümün oldukça iyi sonuç verdiği tespit edilmiştir. Bu sebeple, normale yaklaşım göz önünde bulundurularak DiabetesPedigreeFunction ve Age değişkenlerinin logaritmik dönüşümlü halleri ile analize devam edilecektir.

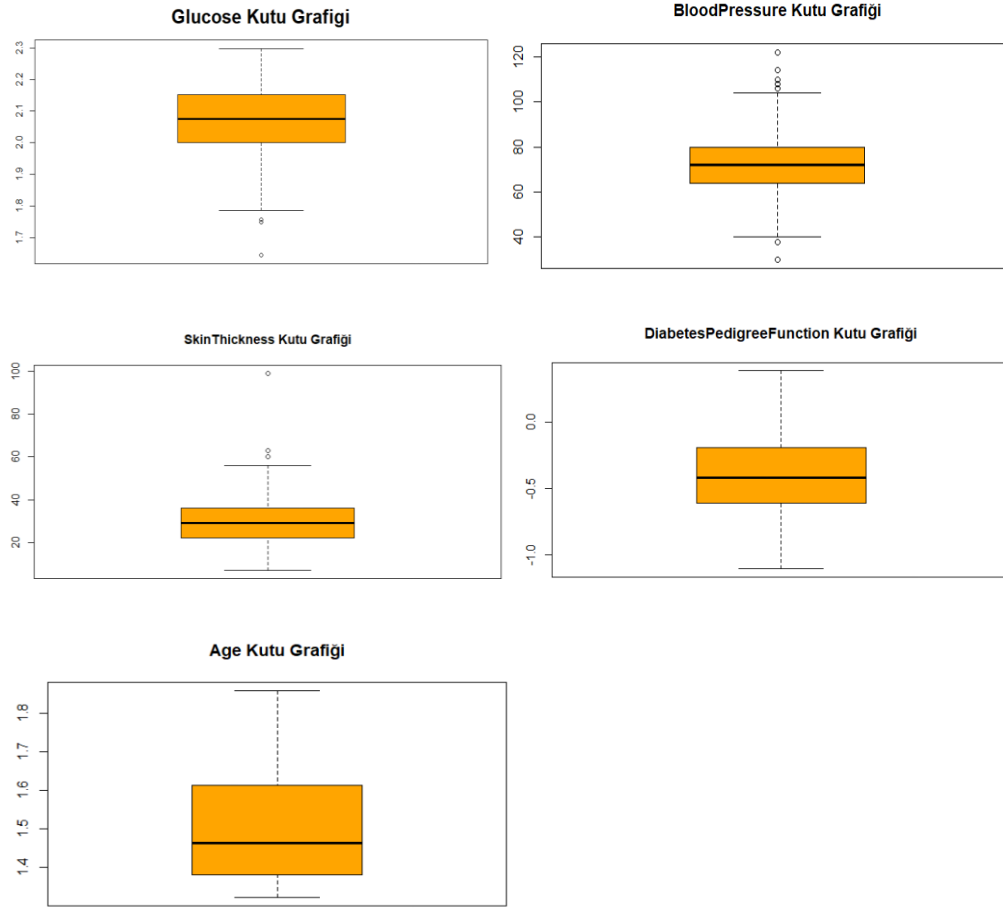
10.5 Sürekli Değişkenlerin Histogram Grafikleri



Şekil 10.5: Sürekli Değişkenlerin Histogram Grafikleri (R Programlama Çıktısı)

- Glucose değişkeninin histogramı incelendiğinde dağılımın, sağ tarafta yoğunlaştığı görülmektedir. Normallik incelemesinden Glucose dağılımının, normal dağıldığı biliniyordu. Histograma bakıldığında da dağılımın, normalliği görülmektedir.
- BloodPressure değişkeninin histogramı incelendiğinde normale yakın bir dağılımı olduğu söylenebilir.
- SkinThickness değişkeninin histogramı incelendiğinde dağılımın, sağa çarpık olduğu görülmektedir. Ancak genel anlamda dağılım, normale yakın durmaktadır.
- DiabetesPedigreeFunction değişkeninin histogramı incelendiğinde dağılımın normale oldukça yaklaştığı görülmektedir. Çift tepelidir.
- Age değişkeninin histogramı incelendiğinde dağılımın, basık ve sağa çarpık olduğu görülmektedir.

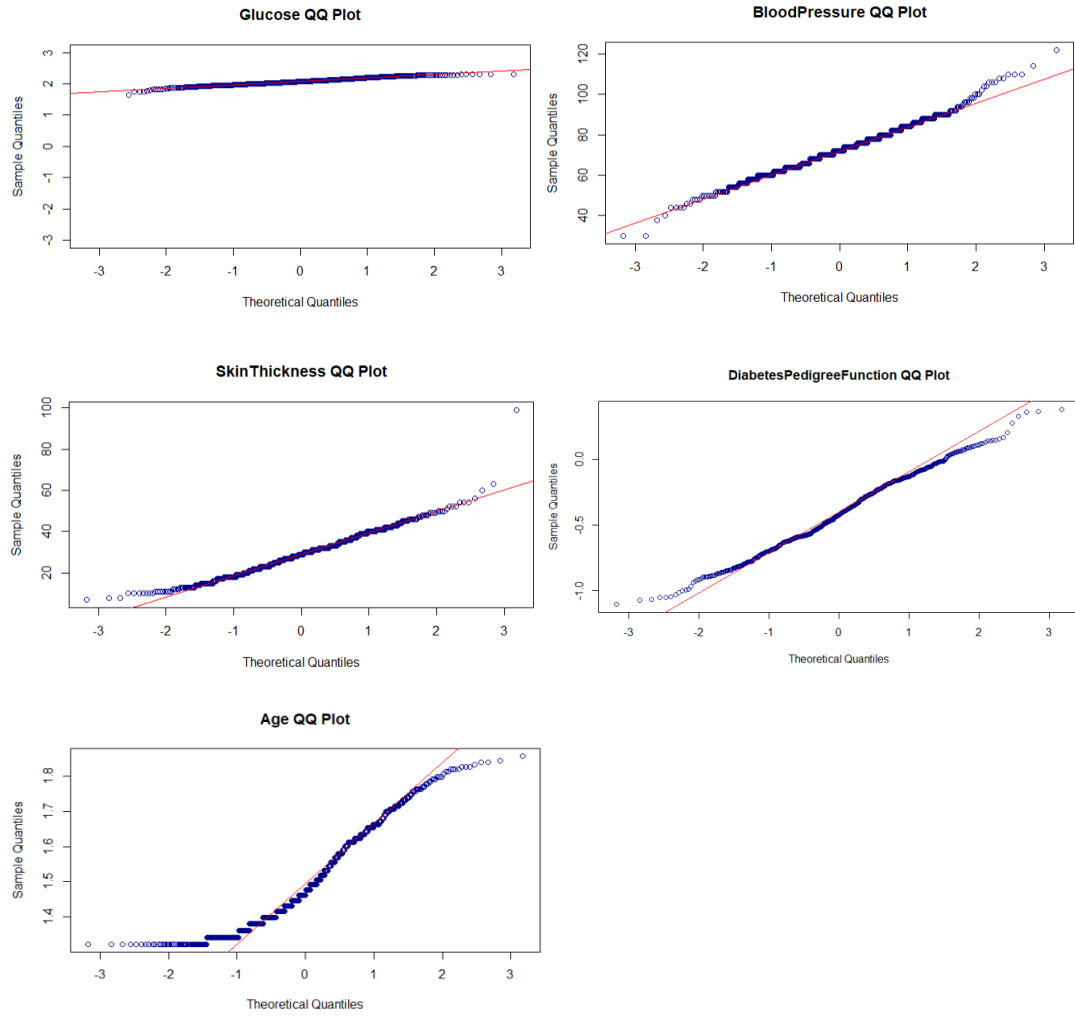
10.6 Sürekli Değişkenlerin Kutu Grafikleri



Şekil 10.6: Sürekli Değişkenlerin Kutu Grafikleri (R Programlama Çıktısı)

- Glucose değişkeninin normallik varsayımını dönüşüm ile sağladığı bilinmekteydi. Grafik de bu durumu onaylamaktadır. Grafikte 3 tane uç değer görülmektedir.
- BloodPressure dağılımı, normale oldukça yakın durmaktadır. Uç değerler olmasına karşın, normale yakın bir dağılıma sahip olduğu görülmektedir.
- SkinThickness dağılımı, normale oldukça yakın durmaktadır. 3 adet uç değer göze çarpmaktadır.
- DiabetesPedigreeFunction dağılımı incelendiğinde, değişkenin normale oldukça yaklaştığı görülmektedir.
- Age değişkeni incelendiğinde, değişkenin sağa çarpık olduğu görülmektedir. Ancak dönüşüm ile normale oldukça yaklaştığı, histogram grafiğinden de görülmektedir.

10.7 Sürekli Değişkenlerin Q-Q Grafikleri

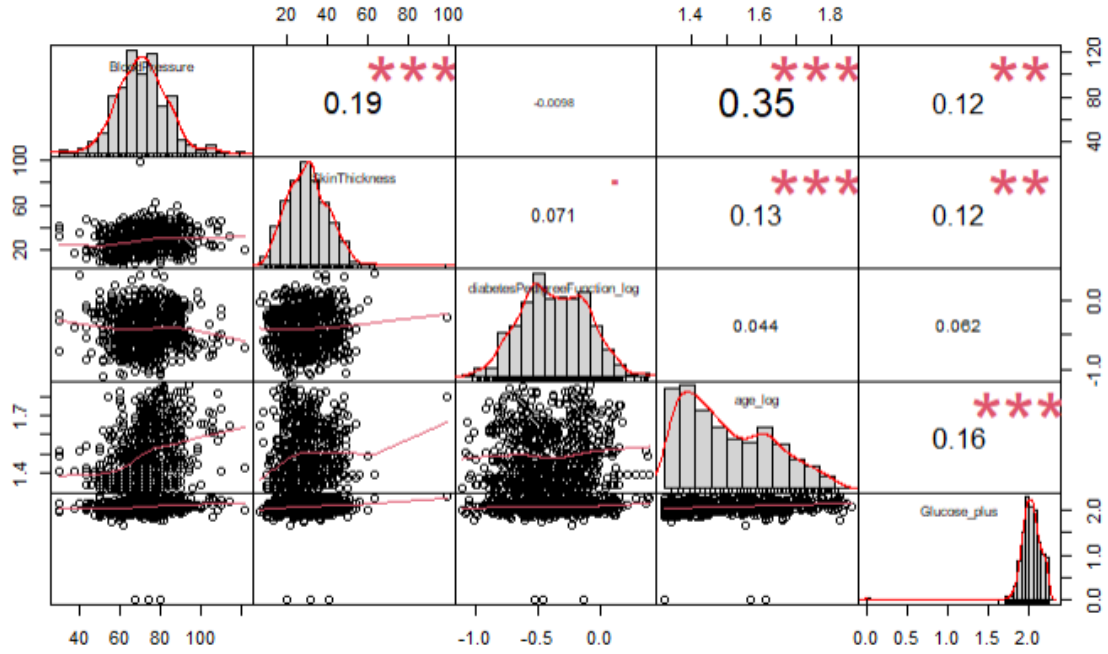


Şekil 10.7: Sürekli Değişkenlerin Q-Q Grafikleri (R Programlama Çıktısı)

- Glucose değişkeninin Q-Q grafiği, normal dağılıma sahip olduğunu göstermektedir.
- BloodPressure değişkeninde normale oldukça yakın bir dağılım söz konusudur. Uç değerlerden kaynaklı sapmalar görülmektedir.
- SkinThickness değişkeninde de uç değerlerden kaynaklı ufak sapmalar dışında normale oldukça yakın bir dağılım söz konusudur.
- DiabetesPedigreeFunction değişkeninde de uç değerlerin sapmaları dışında normale oldukça yakın bir grafik görülmektedir.
- Age değişkeni normal dağılıma çok uzak olmamakla birlikte, uç değer varlığı görülmektedir.

10.8 Sürekli Değişkenlerin Korelasyon Analizi

Glucose, BloodPressure, SkinThickness, DiabetesPedigreeFunction ve Age değişkenlerinin neredeyse normal dağıldıkları Histogram, Kutu Grafiği ve Q-Q Grafik ile söylenmişti. Bu sebeple Korelasyon incelemesinde, belirtilen değişkenler için Pearson Korelasyon Katsayısı kullanılacaktır.



Şekil 10.8: Sürekli Değişkenlerin Korelasyon Grafiği (R Programlama Çıktısı)

Korelasyon haritası incelendiğinde, değişkenler arasında çok yüksek ilişki olmadığı söylenebilir. Genel bir inceleme yapıldığında, en yüksek ilişkinin 0.35 değeri ile BloodPressure ve age_log değişkenleri arasında olduğu görülmektedir. Ardından gelen en büyük ilişki 0.19 değeri ile BloodPressure ve SkinThickness değişkenlerine aittir.

10.9 Ki-Kare Bağımsızlık Analizi

- İnsulin - BMI Analizi

İnsulin * BMI Crosstabulation

Count		BMI				Total
		İdeal	Kilolu	Morbid Obez	Obez	
İnsulin	Diyabet	2	8	20	50	80
	Gizli Şeker	6	12	9	49	76
	Normal	81	130	60	243	514
Total		89	150	89	342	670

Şekil 10.9.1: İnsulin ve BMI Değişkenleri Çapraz Tablosu (SPSS Çıktısı)

İnsulin değişkeni “Diyabet”, “Gizli Şeker” ve “Normal” olmak üzere 3 kategoriye; BMI değişkeni ise “İdeal”, “Kilolu”, “Obez” ve “Morbid Obez” olmak üzere 4 kategoriye ayrılmaktadır.

İnsulin durumu “Diyabet” olan toplam 80 kişi vardır. 80 kişinin 2’sinin BMI değişkeni “İdeal”, 8’inin “Kilolu”, 20’sinin “Morbid Obez” ve 50’sinin “Obez” dir. İnsulin durumu “Gizli Şeker” olan toplam 76 kişi vardır. 76 kişinin 6’sının BMI değişkeni “İdeal”, 12’sinin “Kilolu”, 9’unun “Morbid Obez” ve 49’unun “Obez” dir. İnsulin durumu “Normal” olan toplam 514 kişi vardır. 514 kişinin 81’inin BMI değişkeni “İdeal”, 130’unun “Kilolu”, 60’ının “Morbid Obez” ve 243’ünün “Obez” dir.

Ki-Kare Analizi ile $\alpha = 0.05$ düzeyinde, veri setinde yer alan İnsulin ve BMI değişkenlerinin bağımsızlıkları test edilecektir.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	35,494 ^a	6	,000
Likelihood Ratio	38,767	6	,000
N of Valid Cases	670		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,10.

Şekil 10.9.2: Kİ-Kare Testi (SPSS Çıktısı)

Hipotezler:

H₀: Insulin ve BMI değişkenleri bağımsızdır.

H_a: Insulin ve BMI değişkenleri bağımlıdır.

Yorum: Significance değeri $\alpha = 0.05$ değerinden küçük olduğu görülmektedir. Bu nedenle H₀, yani “Insulin ve BMI değişkenleri bağımsızdır.” hipotezi %95 güven düzeyi ile reddedilir.

- Insulin – Outcome Analizi

Insulin * Outcome Crosstabulation				
Count		Outcome		Total
		Diyabeti Var	Diyabeti Yok	
Insulin	Diyabet	47	33	80
	Gizli Şeker	40	36	76
	Normal	181	333	514
Total		268	402	670

Şekil 10.9.3: Insulin ve Outcome Değişkenleri Çapraz Tablosu (SPSS Çıktısı)

Insulin değişkeni “Diyabet”, “Gizli Şeker” ve “Normal” olmak üzere 3 kategoriye; Outcome değişkeni ise “Diyabeti Var” ve “Diyabeti Yok” olmak üzere 2 kategoriye ayrılmaktadır.

Insulin durumu “Diyabet” olan toplam 80 kişi vardır. 80 kişinin 47’sinin Outcome değişkeni “Diyabeti Var” ve 33’ünün “Diyabeti Yok” tur. Buradan anlaşılabacağı üzere, Insulin değeri Diyabet sınırları içinde olan 33 gözleme diyabet tanısı konulmamıştır. Insulin durumu “Gizli Şeker” olan toplam 76 kişi vardır. 76 kişinin 40’ının Outcome değişkeni “Diyabeti Var” ve 36’sının “Diyabeti Yok” tur. Insulin durumu “Normal” olan toplam 514 kişi vardır. 514 kişinin 181’inin Outcome değişkeni “Diyabeti Var” ve 333’ünün “Diyabeti Yok” tur.

Ki-Kare Analizi ile $\alpha = 0.05$ düzeyinde, veri setinde yer alan Insulin ve Outcome değişkenlerinin bağımsızlıkları test edilecektir.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	21,677 ^a	2	,000
Likelihood Ratio	21,320	2	,000
N of Valid Cases	670		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 30,40.

Şekil 10.9.4: Kİ-Kare Testi (SPSS Çıktısı)

Hipotezler:

H₀: Insulin ve Outcome değişkenleri bağımsızdır.

H_a: Insulin ve Outcome değişkenleri bağımlıdır.

Yorum: Significance değeri $\alpha = 0.05$ değerinden küçük olduğu görülmektedir. Bu nedenle H₀, yani “Insulin ve Outcome değişkenleri bağımsızdır.” hipotezi %95 güven düzeyi ile reddedilir.

- Outcome – BMI Analizi

Outcome * BMI Crosstabulation

Count		BMI				Total
		İdeal	Kilolu	Morbid Obez	Obez	
Outcome	Diyabeti Var	7	40	56	165	268
	Diyabeti Yok	82	110	33	177	402
Total		89	150	89	342	670

Şekil 10.9.5: Insulin ve Outcome Değişkenleri Çapraz Tablosu (SPSS Çıktısı)

Outcome değişkeni “Diyabeti Var” ve “Diyabeti Yok” olmak üzere 2 kategoriye; BMI değişkeni ise “İdeal”, “Kilolu”, ”Obez” ve “Morbid Obez” olmak üzere 4 kategoriye ayrılmaktadır. Outcome durumu “Diyabeti Var” olan toplam 268 kişi vardır.

268 kişinin 7’sinin BMI değişkeni “İdeal”, 40’ının “Kilolu”, 56’sının “Morbid Obez” ve 165’inin “Obez” dir. Outcome durumu “Diyabeti Yok” olan toplam 402 kişi vardır. 402 kişinin 82’sinin BMI değişkeni “İdeal”, 110’unun “Kilolu”, 33’ünün “Morbid Obez” ve 177’sinin “Obez” dir.

Ki-Kare Analizi ile $\alpha = 0.05$ düzeyinde, veri setinde yer alan Insulin ve Outcome değişkenlerinin bağımsızlıkları test edilecektir.

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	78,577 ^a	3	,000
Likelihood Ratio	87,769	3	,000
N of Valid Cases	670		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 35,60.

Şekil 10.9.6: Kİ-Kare Testi (SPSS Çıktısı)

Hipotezler:

H_0 : Outcome ve BMI değişkenleri bağımsızdır.

H_a : Outcome ve BMI değişkenleri bağımlıdır.

Yorum: Significance değeri $\alpha = 0.05$ değerinden küçük olduğu görülmektedir. Bu nedenle H_0 , yani “Outcome ve BMI değişkenleri bağımsızdır.” hipotezi %95 güven düzeyi ile reddedilir.

10.10 T Testi ve Varyans Analizi

- SkinThickness – Outcome Değişkenleri Testi

Group Statistics					
	Outcome	N	Mean	Std. Deviation	Std. Error Mean
SkinThickness	Diyabeti Yok	402	26,74	10,021	,500
	Diyabeti Var	268	32,94	9,989	,610

Şekil 10.10.1: SkinThickness – Outcome Değişkenleri Grup İstatistikleri (SPSS Çıktısı)

Tabloda ilgili değişkenlerin bazı tanımlayıcı istatistikleri verilmiştir. Diyabeti olmayan gözlemlerin cilt kalınlık ortalamaları 26.74 ± 10.021 ve diyabeti olan gözlemlerin cilt kalınlık ortalamaları 32.94 ± 9.989 olarak bulunmuştur.

Independent Samples Test									
Levene's Test for Equality of Variances				t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
SkinThickness	Equal variances assumed	2,063	,151	-7,851	668	,000	-6,197	,789	Lower -7,746 Upper -4,647
	Equal variances not assumed			-7,856	573,583	,000	-6,197	,789	Lower -7,746 Upper -4,647

Şekil 10.10.2: SkinThickness – Outcome Bağımsız Örneklem Testi (SPSS Çıktısı)

Levene's Test

Gruplar arası varyans eşitliği, $\alpha = 0.05$ düzeyinde test edilecektir.

Hipotezler:

H_0 : Gruplar arası varyans eşittir.

H_a : Grup arası varyans eşit değildir.

Yorum: Significance değeri $0.151 > 0.05$ olduğu için H_0 , “Gruplar arası varyans eşittir.” hipotezi %95 güven düzeyi ile reddedilemez.

T Testi

Diyabete göre cilt kalınlığı ortalamaları arasında fark olup olmadığı durumu, $\alpha = 0.05$ düzeyinde test edilecektir.

Hipotezler:

H_0 : Diyabete göre cilt kalınlığı ortalamaları arasında fark yoktur.

H_a : Diyabete göre cilt kalınlığı ortalamaları arasında fark vardır.

Yorum: Significance değeri $0.000 > 0.05$ olduğu için H_0 , “Diyabete göre cilt kalınlığı ortalamaları arasında fark yoktur.” hipotezi %95 güven düzeyi ile reddedilir.

- Glucose – Outcome Değişkenleri Testi

Group Statistics					
	Outcome	N	Mean	Std. Deviation	Std. Error Mean
glucose_log	Diyabeti Yok	401	2,031588007	,0954475521	,0047664233
	Diyabeti Var	266	2,143585906	,0928703577	,0056942470

Şekil 10.10.3: Glucose – Outcome Değişkenleri Grup İstatistikleri (SPSS Çıktısı)

Diyabeti olmayan gözlemlerin glukoz değer ortalamaları 2.031 ± 0.954 ve diyabeti olan gözlemlerin glukoz değer ortalamaları 2.143 ± 0.092 olarak bulunmuştur.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
glucose_log	Equal variances assumed	,785	,376	-14,999	665	,000	-,111997899	,0074671558	-,126659941	-,097335857
	Equal variances not assumed			-15,082	578,348	,000	-,111997899	,0074258494	-,126582819	-,097412979

Şekil 10.10.4: Glucose – Outcome Bağımsız Örneklem Testi (SPSS Çıktısı)

Levene's Test

Gruplar arası varyans eşitliği, $\alpha = 0.05$ düzeyinde test edilecektir.

Hipotezler:

H_0 : Gruplar arası varyans eşittir.

H_a : Grup arası varyans eşit değildir.

Yorum: Significance değeri $0.376 > 0.05$ olduğu için H_0 , “Gruplar arası varyans eşittir.” hipotezi %95 güven düzeyi ile reddedilemez.

T Testi

Diyabete göre glukoz değer ortalamaları arasında fark olup olmadığı durumu, $\alpha = 0.05$ düzeyinde test edilecektir.

Hipotezler:

H_0 : Diyabete göre glukoz değer ortalamaları arasında fark yoktur.

H_a : Diyabete göre glukoz değer ortalamaları arasında fark vardır.

Yorum: Significance değeri $0.000 > 0.05$ olduğu için H_0 , “Diyabete göre glukoz değer ortalamaları arasında fark yoktur.” hipotezi %95 güven düzeyi ile reddedilir.

10.11 Varyans Analizi

- BMI – DiabetesPedigreeFunction Değişken İncelemesi

Descriptives								
diabetesPedigreeFunction_log								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
İdeal	87	-,476381602	,2713258219	,0290891815	-,534208981	-,418554223	-1,07572071	,1690863575
Kilolu	153	-,446234145	,2828342914	,0228657972	-,491409961	-,401058328	-1,10790540	,2041199827
Obez	342	-,405335310	,2696061723	,0145786321	-,434010679	-,376659940	-1,07058107	,3671694885
Morbid Obez	88	-,326342786	,2944856980	,0313922809	-,388738343	-,263947228	-1,05060999	,3838153660
Total	670	-,413525171	,2788532455	,0107730407	-,434678212	-,392372130	-1,10790540	,3838153660

Şekil 10.11.1: BMI – DiabetesPedigreeFunction Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Tabloda DiabetesPedigreeFunction değişkeninin BMI değişkeni tiplerine göre tanımlayıcı istatistikleri bulunmaktadır. BMI değeri “İdeal” olan bireylerin aile geçmişlerinde diyabet görülme ortalaması -0.476 ± 0.271 , “Kilolu” olan bireylerin aile geçmişlerinde diyabet görülme ortalaması -0.446 ± 0.282 , “Obez” olan bireylerin aile geçmişlerinde diyabet görülme ortalaması -0.405 ± 0.269 ve “Morbid Obez” olan bireylerin aile geçmişlerinde diyabet görülme ortalaması -0.413 ± 0.278 olarak bulunmuştur.

Varyans homojenliği, $\alpha = 0.05$ düzeyinde test edilecektir.

Test of Homogeneity of Variances					
		Levene Statistic	df1	df2	Sig.
diabetesPedigreeFunction_log	Based on Mean	,490	3	666	,689
	Based on Median	,447	3	666	,719
	Based on Median and with adjusted df	,447	3	661,670	,719
	Based on trimmed mean	,492	3	666	,688

Şekil 10.11.2: Varyans Homojenliği Testi (SPSS Çıktısı)

Hipotezler:

H_0 : Grup varyansları homojendir.

H_a : Grup varyansları homojen değildir.

Yorum: Significance değeri > 0.05 olduğu için H_0 , “Grup varyansları homojendir.” hipotezi %95 güven düzeyi ile reddedilemez.

BMI değerlerine göre DiabetesPedigreeFunction değerlerinde farklılık olup olmadığı, $\alpha = 0.05$ düzeyinde test edilecektir.

ANOVA

diabetesPedigreeFunction_log					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,199	3	,400	5,238	,001
Within Groups	50,822	666	,076		
Total	52,021	669			

Şekil 10.11.3: ANOVA (SPSS Çıktısı)

Hipotezler:

H_0 : BMI değerlerine göre DiabetesPedigreeFunction ortalamaları arasında fark yoktur.

H_a : BMI değerlerine göre DiabetesPedigreeFunction ortalamaları arasında fark vardır.

Yorum: Significance değeri $0.001 < 0.05$ olduğu için H_0 , “BMI değerlerine göre DiabetesPedigreeFunction ortalamaları arasında fark yoktur.” hipotezi %95 güven düzeyi ile reddedilir.

Post Hoc Testi’nde varyans homojenliği sağlandığı için Tukey kullanılacaktır.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: diabetesPedigreeFunction_log

Tukey HSD

(I) BMI	(J) BMI	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
İdeal	Kilolu	-,030147458	,0370926159	,848	-,125680563	,0653856478
	Obez	-,071046292	,0331698227	,141	-,156476130	,0143835456
	Morbid Obez	-,150038817*	,0417643037	,002	-,257603989	-,042473644
Kilolu	İdeal	,0301474576	,0370926159	,848	-,065385648	,1256805629
	Obez	-,040898835	,0268677324	,425	-,110097457	,0282997874
	Morbid Obez	-,119891359*	,0369580163	,007	-,215077799	-,024704919
Obez	İdeal	,0710462923	,0331698227	,141	-,014383546	,1564761303
	Kilolu	,0408988348	,0268677324	,425	-,028299787	,1100974569
	Morbid Obez	-,078992524	,0330192361	,080	-,164034522	,0060494737
Morbid Obez	İdeal	,150038817*	,0417643037	,002	,0424736441	,2576039890
	Kilolu	,119891359*	,0369580163	,007	,0247049187	,2150777992
	Obez	,0789925242	,0330192361	,080	-,006049474	,1640345221

*. The mean difference is significant at the 0.05 level.

Şekil 10.11.4: Post Hoc Testi (SPSS Çıktısı)

“İdeal” BMI’a sahip bireylerin aile geçmişlerinde diyabet görülme ortalaması “Morbid Obez” BMI’ına sahip olan bireylerin aile geçmişlerinde diyabet görülme ortalamasından -0.15 birim daha azdır ve bu fark anlamlıdır.

“Kilolu” BMI’a sahip bireylerin aile geçmişlerinde diyabet görülme ortalaması “Morbid Obez” BMI’ına sahip olan bireylerin aile geçmişlerinde diyabet görülme ortalamasından -0.119 birim daha azdır ve bu fark anlamlıdır.

- İnsulin – Glucose Değişken İncelemesi

Descriptives								
glucose_log								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
Normal	511	2,056488446	,1074381850	,0047527855	2,047150999	2,065825894	1,643452676	2,298853076
Diyabet	80	2,169824988	,0835956639	,0093462794	2,151221685	2,188428292	1,959041392	2,296665190
Gizli Şeker	76	2,110645354	,0821942493	,0094283270	2,091863163	2,129427544	1,968482949	2,290034611
Total	667	2,076252837	,1091575895	,0042265987	2,067953773	2,084551900	1,643452676	2,298853076

Şekil 10.11.5: İnsulin – Glucose Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Tabloda Glucose değişkeninin İnsulin değişkeni tiplerine göre tanımlayıcı istatistikleri bulunmaktadır. İnsulin değeri “Normal” olan bireylerin glukoz değer ortalaması 2.056 ± 0.107 , “Diyabet” olan bireylerin glukoz değer ortalaması 2.169 ± 0.083 ve “Gizli şeker” olan bireylerin glukoz değer ortalaması 2.110 ± 0.821 olarak bulunmuştur.

Varyans homojenliği, $\alpha = 0.05$ düzeyinde test edilecektir.

Test of Homogeneity of Variances					
		Levene Statistic	df1	df2	Sig.
glucose_log	Based on Mean	4,554	2	664	,011
	Based on Median	4,528	2	664	,011
	Based on Median and with adjusted df	4,528	2	630,189	,011
	Based on trimmed mean	4,633	2	664	,010

Şekil 10.11.6: Varyans Homojenliği Testi (SPSS Çıktısı)

Hipotezler:

H₀: Grup varyansları homojendir.

H_a: Grup varyansları homojen değildir.

Yorum: Significance değeri < 0.05 olduğu için H₀, “Grup varyansları homojendir.” hipotezi %95 güven düzeyi ile reddedilir.

Insulin değerlerine göre Glucose değerleri ortalamaları arasında farklılık olup olmadığı, $\alpha = 0.05$ düzeyinde test edilecektir.

Robust Tests of Equality of Means

glucose_log

	Statistic ^a	df1	df2	Sig.
Welch	62,594	2	146,879	,000
Brown-Forsythe	66,728	2	226,713	,000

a. Asymptotically F distributed.

Şekil 10.11.7: Robust Test (SPSS Çıktısı)

Hipotezler:

H₀: Insulin durumuna göre Glucose değer ortalamaları arasında fark yoktur.

H_a: Insulin durumuna göre Glucose değer ortalamaları arasında fark vardır.

Yorum: Significance değeri < 0.05 olduğu için H₀, “Insulin durumuna göre Glucose değer ortalamaları arasında fark yoktur.” hipotezi %95 güven düzeyi ile reddedilir.

Post Hoc Testinde varyans homojenliği sağlanmadığı için Tamhane kullanılacaktır.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: glucose_log

Tamhane

(I) Insulin	(J) Insulin	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Normal	Diyabet	-,113336542*	,0104853187	,000	-,138715787	-,087957297
	Gizli Şeker	-,054156908*	,0105585188	,000	-,079734331	-,028579484
Diyabet	Normal	,113336542*	,0104853187	,000	,0879572972	,1387157871
	Gizli Şeker	,059179635*	,0132757782	,000	,0271316826	,0912275864
Gizli Şeker	Normal	,054156908*	,0105585188	,000	,0285794837	,0797343315
	Diyabet	-,059179635*	,0132757782	,000	-,091227586	-,027131683

*. The mean difference is significant at the 0.05 level.

Şekil 10.11.8: Post Hoc Testi (SPSS Çıktısı)

“Normal” Insulin değerine sahip bireylerin Glucose ortalaması, “Diyabet” Insulin değerine sahip olan bireylerin Glucose ortalamasından -0.113 birim daha azdır ve bu fark anlamlıdır.

“Normal” Insulin değerine sahip bireylerin Glucose ortalaması, “Gizli Şeker” Insulin değerine sahip olan bireylerin Glucose ortalamasından -0.054 birim daha azdır ve bu fark anlamlıdır.

- BMI – SkinThickness Değişken İncelemesi

Descriptives

SkinThickness

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
İdeal	87	18,68	6,309	,676	17,33	20,02	8	38
Kilolu	153	23,39	6,932	,560	22,28	24,49	7	39
Obez	342	31,96	9,482	,513	30,95	32,96	7	99
Morbid Obez	88	39,17	7,990	,852	37,48	40,86	23	63
Total	670	29,22	10,452	,404	28,43	30,02	7	99

Şekil 10.11.9: BMI – DiabetesPedigreeFunction Değişkenleri Tanımlayıcı İstatistikleri (SPSS Çıktısı)

Tabloda SkinThickness değişkeninin BMI değişkeni tiplerine göre tanımlayıcı istatistikleri bulunmaktadır. BMI değeri “İdeal” olan bireylerin cilt kalınlık ortalaması 18.68 ± 6.309 , “Kilolu” olan cilt kalınlık ortalaması 23.39 ± 6.932 , “Obez” olan bireylerin cilt kalınlık ortalaması 31.96 ± 9.482 ve “Morbid Obez” olan bireylerin cilt kalınlık ortalaması 39.17 ± 7.990 olarak bulunmuştur.

Varyans homojenliği, $\alpha = 0.05$ düzeyinde test edilecektir.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
SkinThickness	Based on Mean	6,300	3	666	,000
	Based on Median	6,768	3	666	,000
	Based on Median and with adjusted df	6,768	3	608,748	,000
	Based on trimmed mean	6,537	3	666	,000

Şekil 10.11.10: Varyans Homojenliği Testi (SPSS Çıktısı)

Hipotezler:

H_0 : Grup varyansları homojendir.

H_a : Grup varyansları homojen değildir.

Yorum: Significance değeri < 0.05 olduğu için H_0 , “Grup varyansları homojendir.” hipotezi %95 güven düzeyi ile reddedilir.

BMI değerlerine göre SkinThickness ortalamaları arasında farklılık olup olmadığı, $\alpha = 0.05$ düzeyinde test edilecektir.

Robust Tests of Equality of Means

SkinThickness				
	Statistic ^a	df1	df2	Sig.
Welch	161,534	3	245,048	,000
Brown-Forsythe	152,758	3	457,688	,000

a. Asymptotically F distributed.

Şekil 10.11.11: Robust Test (SPSS Çıktısı)

Hipotezler:

H₀: BMI durumuna göre SkinThickness ortalamaları arasında fark yoktur.

H_a: BMI durumuna göre SkinThickness ortalamaları arasında fark vardır.

Yorum: Significance değeri < 0.05 olduğu için H₀, “BMI durumuna göre SkinThickness ortalamaları arasında fark yoktur.” hipotezi %95 güven düzeyi ile reddedilir.

Post Hoc Testi’nde varyans homojenliği sağlanmadığı için Tamhane kullanılacaktır.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: SkinThickness

Tamhane

(I) BMI	(J) BMI	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
İdeal	Kilolu	-4,707*	,878	,000	-7,04	-2,37
	Obez	-13,278*	,849	,000	-15,53	-11,02
	Morbid Obez	-20,492*	1,088	,000	-23,39	-17,60
Kilolu	İdeal	4,707*	,878	,000	2,37	7,04
	Obez	-8,571*	,760	,000	-10,58	-6,56
	Morbid Obez	-15,785*	1,020	,000	-18,50	-13,07
Obez	İdeal	13,278*	,849	,000	11,02	15,53
	Kilolu	8,571*	,760	,000	6,56	10,58
	Morbid Obez	-7,214*	,994	,000	-9,86	-4,57
Morbid Obez	İdeal	20,492*	1,088	,000	17,60	23,39
	Kilolu	15,785*	1,020	,000	13,07	18,50
	Obez	7,214*	,994	,000	4,57	9,86

*. The mean difference is significant at the 0.05 level.

Şekil 10.11.12: Post Hoc Testi (SPSS Çıktısı)

“İdeal” BMI’a sahip bireylerin aile geçmişlerinde diyabet görülme ortalaması “Kilolu” BMI’ına sahip olan bireylerin aile geçmişlerinde diyabet görülme ortalamasından -4.707 birim daha azdır ve bu fark anlamlıdır.

“İdeal” BMI’a sahip bireylerin aile geçmişlerinde diyabet görülme ortalaması “Obez” BMI’ına sahip olan bireylerin aile geçmişlerinde diyabet görülme ortalamasından -13.278 birim daha azdır ve bu fark anlamlıdır.

“İdeal” BMI’a sahip bireylerin aile geçmişlerinde diyabet görülme ortalaması “Morbid Obez” BMI’ına sahip olan bireylerin aile geçmişlerinde diyabet görülme ortalamasından -20.492 birim daha azdır ve bu fark anlamlıdır.

10.12 Lojistik Regresyon Analizi

Değişkenlerin modele katkısı, $\alpha = 0.05$ düzeyinde test edilecektir.

Omnibus Test ^a		
Likelihood Ratio Chi- Square	df	Sig.
290,021	11	,000
Dependent Variable: Outcome Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log		
a. Compares the fitted model against the intercept-only model.		

Şekil 10.12.1: Omnibus Testi (SPSS Çıktısı)

Hipotezler:

H₀: Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamamaktadır.

H_a: Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamaktadır.

Yorum: Significance değeri = 0.000 < $\alpha = 0.05$ olduğu tüm değişkenler için görülmektedir. Bu nedenle H₀, yani “Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamamaktadır.” hipotezi %95 güven düzeyi ile reddedilir.

Değişkenlerin, modele katkı sağlama oranlarının anlamlılıkları $\alpha = 0.05$ düzeyinde test edilecektir.

Tests of Model Effects

Source	Wald Chi-Square	Type III	
		df	Sig.
(Intercept)	105,024	1	,000
Insulin	3,476	2	,176
BMI	20,197	3	,000
Pregnancies	10,667	1	,001
BloodPressure	,213	1	,645
SkinThickness	3,224	1	,073
glucose_log	88,446	1	,000
diabetesPedigreeFunction_log	12,926	1	,000
age_log	3,282	1	,070

Dependent Variable: Outcome

Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

Şekil 10.12.2: Bağımsız Değişkenlerin Modele Etkileri (SPSS Çıktısı)

Hipotezler:

H₀: Bağımsız değişkenlerin modele katkıları anlamsızdır.

H_a: Bağımsız değişkenlerin modele katkıları anlamlıdır.

Yorum: Insulin, BloodPressure, SkinThickness ve Age değişkenleri için Significance değerleri $> \alpha = 0.05$ olduğu görülmektedir. Bu nedenle H₀, yani “Bağımsız değişkenlerin modele katkıları anlamsızdır.” hipotezi %95 güven düzeyi ile reddedilemez.

Modelin verilere uyumluluğu Ki-Kare ile test edilecektir.

Goodness of Fit^a

	Value	df	Value/df
Deviance	607,124	655	,927
Scaled Deviance	607,124	655	
Pearson Chi-Square	607,762	655	,928
Scaled Pearson Chi-Square	607,762	655	
Log Likelihood ^b	-303,562		
Akaike's Information Criterion (AIC)	631,124		
Finite Sample Corrected AIC (AICC)	631,601		
Bayesian Information Criterion (BIC)	685,157		
Consistent AIC (CAIC)	697,157		

Dependent Variable: Outcome

Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Şekil 10.12.3: Uyumluluk Testi (SPSS Çıktısı)

Hipotezler:

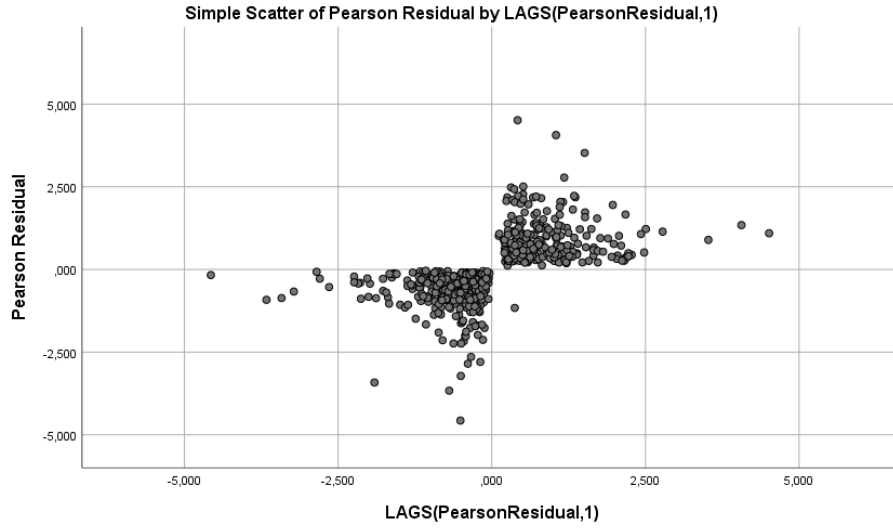
H₀: Model, veriler ile uyumludur.

H_a: Model, veriler ile uyumlu değildir.

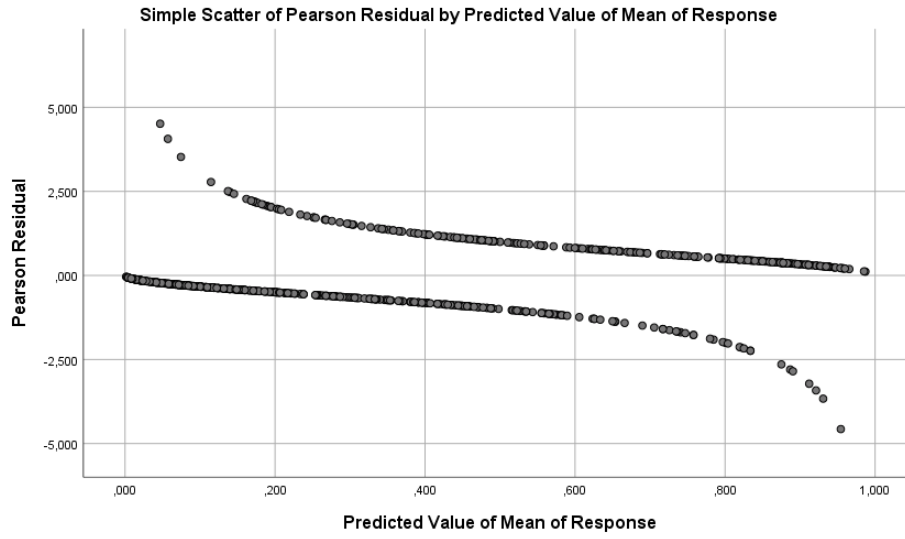
Yorum: $X^2_{655;0.05} > 607,762$ olduğu için H₀, yani “Model, veriler ile uyumludur.” hipotezi %95 güven düzeyi ile reddedilemez.

10.13 Varsayım İncelemesi

GGraph



Şekil 10.13.1: Otokorelasyon İncelemesi (SPSS Çıktısı)



Şekil 10.13.2: Değişen Varyanslılık İncelemesi (SPSS Çıktısı)

Şekil 10.13.1 incelendiğinde otokorelasyon, şekil 10.13.2 incelendiğinde ise rastgelelik olmaması durumları gözlemlenmektedir. Bu varsayım bozuklukları için “Sağlam Kestirim” ile model yeniden kurulup analize eklenmiştir.

10.14 Çoklu Doğrusal Bağlantı Sorununun İncelenmesi

Coefficients ^a		
Model	Collinearity Statistics	
	Tolerance	VIF
1		
(Constant)		
Pregnancies	,663	1,507
BloodPressure	,810	1,235
SkinThickness	,631	1,584
Insulin	,897	1,115
BMI	,606	1,651
glucose_log	,818	1,222
diabetesPedigreeFunction_log	,952	1,051
age_log	,585	1,709

a. Dependent Variable: Outcome

Şekil 10.14: VIF Değerleri (SPSS Çıktısı)

Çoklu doğrusal bağlantının var olması için VIF değerlerinin 5'ten büyük olması gerekmektedir. Değişkenlerin VIF değerleri incelendiğinde hepsinin 5'ten küçük olduğu görülmektedir. Bu sebeple çoklu doğrusal bağlantı yoktur.

10.15 Sağlam Kestirim Yöntemi ile Lojistik Regresyon Analizi

Değişkenlerin modele katkısı, $\alpha = 0.05$ düzeyinde test edilecektir.

Omnibus Test ^a		
Likelihood Ratio Chi-Square	df	Sig.
290,021	11	,000

Dependent Variable: Outcome
Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

a. Compares the fitted model against the intercept-only model.

Şekil 10.15.1: Omnibus Testi (SPSS Çıktısı)

Hipotezler:

H₀: Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamamaktadır.

H_a: Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamaktadır.

Yorum: Significance değeri = 0.000 < α = 0.05 olduğu tüm değişkenler için görülmektedir. Bu nedenle H₀, yani “Modele eklenen bağımsız değişkenler, modele anlamlı katkı sağlamamaktadır.” hipotezi %95 güven düzeyi ile reddedilir.

Değişkenlerin, modele katkı sağlama oranlarının anlamlılıkları α = 0.05 düzeyinde test edilecektir.

Tests of Model Effects

Source	Wald Chi-Square	Type III	
		df	Sig.
(Intercept)	101,049	1	,000
Insulin	3,597	2	,166
BMI	23,740	3	,000
Pregnancies	9,498	1	,002
BloodPressure	,223	1	,637
SkinThickness	3,188	1	,074
glucose_log	87,223	1	,000
diabetesPedigreeFunction_log	12,957	1	,000
age_log	2,836	1	,092

Dependent Variable: Outcome

Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

Şekil 10.15.2: Bağımsız Değişkenlerin Modele Etkileri (SPSS Çıktısı)

Hipotezler:

H₀: Bağımsız değişkenlerin modele katkıları anlamsızdır.

H_a: Bağımsız değişkenlerin modele katkıları anlamlıdır.

Yorum: Insulin, BloodPressure, SkinThickness ve Age değişkenleri için Significance değerleri $> \alpha = 0.05$ olduğu görülmektedir. Bu nedenle H_0 , yani “Bağımsız değişkenlerin modele katkıları anlamsızdır.” hipotezi %95 güven düzeyi ile reddedilemez.

Modelin verilere uyumluluğu test edilecektir.

Goodness of Fit^a			
	Value	df	Value/df
Deviance	607,124	655	,927
Scaled Deviance	607,124	655	
Pearson Chi-Square	607,762	655	,928
Scaled Pearson Chi-Square	607,762	655	
Log Likelihood ^b	-303,562		
Akaike's Information Criterion (AIC)	631,124		
Finite Sample Corrected AIC (AICC)	631,601		
Bayesian Information Criterion (BIC)	685,157		
Consistent AIC (CAIC)	697,157		

Dependent Variable: Outcome

Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

Şekil 10.15.3: Uyumluluk Testi (SPSS Çıktısı)

Hipotezler:

H_0 : Model, veriler ile uyumludur.

H_a : Model, veriler ile uyumlu değildir.

Yorum: $X^2_{655;0.05} > 607,762$ olduğu için H_0 , yani “Model, veriler ile uyumludur.” hipotezi %95 güven düzeyi ile reddedilemez.

10.16 Model Denklemi

Parametre tahminlemesi ve model denklemi:

Parameter Estimates										
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-29,260	2,7821	-34,713	-23,807	110,609	1	,000	1,962E-13	8,405E-16	4,579E-11
[Insulin=2]	,155	,2991	-,431	,742	,270	1	,603	1,168	,650	2,100
[Insulin=1]	-,532	,3203	-1,160	,096	2,758	1	,097	,587	,314	1,101
[Insulin=0]	0 ^a	1	.	.
[BMI=3]	2,196	,5122	1,192	3,200	18,386	1	,000	8,992	3,295	24,541
[BMI=2]	1,860	,4146	1,047	2,672	20,123	1	,000	6,422	2,850	14,473
[BMI=1]	1,158	,4390	,297	2,018	6,953	1	,008	3,182	1,346	7,525
[BMI=0]	0 ^a	1	.	.
Pregnancies	,119	,0386	,043	,195	9,498	1	,002	1,126	1,044	1,215
BloodPressure	-,004	,0089	-,022	,013	,223	1	,637	,996	,979	1,013
SkinThickness	,022	,0124	-,002	,046	3,188	1	,074	1,022	,998	1,048
glucose_log	11,721	1,2550	9,261	14,181	87,223	1	,000	123110,187	10520,464	1440632,077
diabetesPedigreeFunction_log	1,377	,3824	,627	2,126	12,957	1	,000	3,961	1,872	8,382
age_log	1,695	1,0065	-,278	3,668	2,836	1	,092	5,446	,757	39,158
(Scale)	1 ^b

Dependent Variable: Outcome

Model: (Intercept), Insulin, BMI, Pregnancies, BloodPressure, SkinThickness, glucose_log, diabetesPedigreeFunction_log, age_log

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Şekil 10.16: Parametre Tahminleri (SPSS Çıktısı)

Model denklemi:

$$\ln\left(\frac{p}{1-p}\right) = - 29.260 + 0.119 \text{ (Pregnancies)} - 0.004 \text{ (BloodPressure)} + 0.022 \text{ (SkinThickness)} + 11.721 \text{ (Glucose)} + 1.377 \text{ (DiabetesPedigreeFunction)} + 1.695 \text{ (Age)}$$

Yorumlar:

- Diğer değişkenlerin etkisi sabit tutulduğunda Pregnancies değişkenindeki 1 birimlik artış, diyabet hastası olma ihtimalini olmama ihtimaline göre 1.126 kat arttırmaktadır.
- Diğer değişkenlerin etkisi sabit tutulduğunda BloodPressure değişkenindeki 1 birimlik artış, diyabet hastası olma ihtimalini olmama ihtimaline göre 0.996 kat arttırmaktadır.
- Diğer değişkenlerin etkisi sabit tutulduğunda SkinThickness değişkenindeki 1 birimlik artış, diyabet hastası olma ihtimalini olmama ihtimaline göre 1.022 kat arttırmaktadır.

- Diğer değişkenlerin etkisi sabit tutulduğunda Insulin değişkenindeki kategoriler arttıkça (sıralama: Normal, Gizli Şeker, Diyabet), diyabet hastası olma ihtimalini olmama ihtimaline göre artmaktadır.
- Diğer değişkenlerin etkisi sabit tutulduğunda BMI değişkenindeki kategoriler arttıkça (sıralama: İdeal, Kilolu, Obez, Morbid Obez), diyabet hastası olma ihtimalini olmama ihtimaline göre artmaktadır.
- Diğer değişkenlerin etkisi sabit tutulduğunda DiabetesPedigreeFunction değişkenindeki 1 birimlik artış, diyabet hastası olma ihtimalini olmama ihtimaline göre 3,961 kat arttırmaktadır.
- Diğer değişkenlerin etkisi sabit tutulduğunda Age değişkenindeki 1 birimlik artış, diyabet hastası olma ihtimalini olmama ihtimaline göre 5,446 kat arttırmaktadır.

KAYNAKÇA

- [1]: Türkiye Diyabet Vakfı. [<https://www.turkdiab.org/bilgiler.asp?lang=TR&id=73>]. (Erişim: 10.11.2023)
- [2]: Novo Nordisk Türkiye, Diyabet ve Yaşam. *Diyabeti Değiştirmek*. [<https://www.novonordisk.com.tr/content/dam/Turkey/AFFILIATES/www-novonordisk-com-tr/Home/Patients/Documents/DiyabetVeYasam21-02-2014.pdf>]. (Erişim: 10.11.2023)
- [3]: T.C. Sağlık Bakanlığı. (14.08.2022) *Dünya Diyabet Günü (2020)* [<https://sggm.saglik.gov.tr/TR-76887/dunya-diyabet-gunu-2020.html>]. (Erişim: 10.11.2023)
- [4]: International Diabetes Federation. *Facts & Figures* [<https://idf.org/about-diabetes/diabetes-facts-figures/>]. (Erişim: 10.11.2023)
- [5]: Prof. Dr. Mahir Cengiz. (18.11.2022) *Şeker Hastalığı (Diyabet) Nedir? Şeker Hastalığı Belirtileri Nelerdir?* [<https://www.medicalpark.com.tr/seker-hastaligi-diyabet-nedir/hg-1703#:~:text=%C3%96%C4%9F%C3%BCn%20ba%C5%9Flang%C4%B1c%C4%B1ndan%20%20saat%20sonra,verilen%20pre%2Ddiyabet%20d%C3%B6neminin%20g%C3%B6stergesidir>]. (Erişim: 10.11.2023)
- [6]: Zirekoğlu Merve. (25.12.2022) *ÖRNEKLERLE MERKEZİ EĞİLİM ÖLÇÜLERİ* [[ÖRNEKLERLE MERKEZİ EĞİLİM ÖLÇÜLERİ](https://www.medium.com/@merve.zirekoğlu/örneklerle-merkezi-eğilim-ölçüleri) | by Merve Zirekoğlu | Medium]. (Erişim: 12.12.2023)
- [7]: Erar Aydın. (2023). Veri Analizi Ders Notları, MSGSÜ, İstanbul.
- [8]: Yalın Dünya. (19.06.2020) *KUTU GRAFİĞİ (BOXPLOTS)* [<https://yalin-dunya.com/2020/06/19/kutu-grafigi-boxplot/>]. (Erişim: 12.12.2023)
- [9]: Mesut. (23.02.2020) *Histogram Nedir ? -Histogram Konu Anlatımı ve Örnekleri* [<https://www.matematiksaati.com/histogram-nedir-histogram-konu-anlatimi-ve-ornekleri.html>]. (Erişim: 12.12.2023)
- [10]: Moriarty Therese. (03.09.2019) *The Right Way to Make a Pie Chart* [[The Right Way to Make a Pie Chart](https://www.medium.com/@therese.moriarty/the-right-way-to-make-a-pie-chart) | by Therese Moriarty | eyeful | Medium]. (Erişim: 12.12.2023)

- [11]: Varshney Paras. (29.09.2020) *Q-Q Plots Explained* [<https://medium.com/preai/q-q-plots-explained-75f5bc6d68be>]. (Erişim: 12.12.2023)
- [12]: Weitz Darío. (20.02.2020) *Scatter Plots, Why & How* [[Scatter Plots, Why & How. Storytelling, Tips & Warnings | by Darío Weitz | Analytics Vidhya | Medium](https://medium.com/analytics-vidhya/scatter-plots-why-and-how-storytelling-tips-and-warnings-by-dario-weitz-analytics-vidhya-medium)]. (Erişim: 12.12.2023)
- [13]: Dayanıklı A. Sergen. (08.07.2021) *Python'da Normal Dağılmayan Verilerin Dönüştürülme (Transformation) Yöntemleri* [<https://ravenfo.com/2021/07/08/python-veri-donusturme-data-transformation/>]. (Erişim: 12.12.2023)
- [14]: İstatistik Ders Notları, Ekolar. (11.02.2018) *Kİ KARE TESTLERİ* [<http://www.ekolar.com/ki-kare-testleri/>]. (Erişim: 12.12.2023)
- [15]: Yılmaz Elif. (26.08.2019) *Parametrik Testlerden T Testi ile Merhaba* [<https://www.veribilimiokulu.com/parametrik-testlerden-t-testi-ile-merhaba/>]. (Erişim: 13.12.2023)
- [16]: Yılmaz S. Furkan. (21.07.2020) *Veri Bilimi için İstatistik II* [<https://medium.com/@senafurkanyilmaz/veri-bilimi-i%C3%A7in-i%C3%A7in-istatistik-ii-8b2b670cec78>]. (Erişim: 13.12.2023)
- [17]: Erden Caner. (12.04.2020) *Minitab ile ANOVA Testleri* [<https://cerden.medium.com/minitab-ile-anova-testleri-fc18153d7890>]. (Erişim: 13.12.2023)
- [18]: Akca M. Fatih. (22.02.2021) *Lojistik Regresyon Nedir? Nasıl Çalışır?* [<https://mfakca.medium.com/lojistik-regresyon-nedir-nas%C4%B1l-%C3%A7al%C4%B1%C5%9F%C4%B1r-4e1d2951c5c1>]. (Erişim: 13.12.2023)
- [19]: Arc (07.06.2018). *Derivative of the Sigmoid function* [<https://towardsdatascience.com/derivative-of-the-sigmoid-function-536880cf918e>]. (Erişim: 13.12.2023)
- [20]: Diabetes Data Anslysis. (Yıl bilinmiyor). Kaggle. [[DiabetesDataAnslysis \(kaggle.com\)](https://www.kaggle.com/datasets/stone-island/diabetes-data-analysis)]. (Erişim: 16.10.2023)
- [21]: Acıbadem Web ve Yayın Kurulu. (6.12.2023) *Vücut Kitle İndeksi Hesaplama* [<https://www.acibadem.com.tr/ilgi-alani/vucut-kitle-indeksi-hesaplama/>]. (Erişim: 07.12.2023)