

Kalp Yetmezliđi Hastalarında Hayatta Kalma Analizi

İsim Soyisim*

Özet

Her yıl kardiyovasküler hastalık 17 milyon insanın ölümüne sebep olmaktadır. Kardiyovasküler hastalığın başında gelen miyokard enfarktüsü ve kalp yetmezliđi şeklinde kendini göstermektedir. Kalp, kanın vücut için gerekli olduđu kadarki kısmını yeteri kadar pompalayamadığında ya da pompaladıđı ancak bunu yaparken normalden daha fazla yorulduđu zaman kalp yetmezliđi gelişmektedir. 2015 yılında 299 kalp yetmezliđi hastası üzerinden yapılan incelemeler sonucunda, hastanın tedavi süreci ve hayatta kalma şansı değerlendirilmek istenmiştir. Hastanın kardiyak fonksiyonları, ileri evre semptomları ve ilaç kullanımı gibi faktörler tedavi etkinliđi açısından değerlendirilmiştir. Kalp gibi hayati bir organın önemi göz önüne alındığında, kalp yetmezliđini tahmin etmek doktorlar için bir öncelik haline gelmiştir. Bu bağlamda, elektronik sağlık kayıtları (tıbbi kayıtlar olarak da adlandırılır), yalnızca araştırma için deđil, aynı zamanda klinik uygulama için de hasta verileri arasındaki gizli ve açık olmayan ilişkileri ortaya çıkarmaya yönelik yararlı bir bilgi kaynađı olarak düşünülebilir. Bu amaçla, son yıllarda risk faktörlerine ilişkin bilgileri derinleştirmek için farklı koşulları ve demografik özellikleri kapsayan, farklı veri kaynakları ile çok sayıda tarama çalışması yapılmıştır. Klinik olarak, kalp yetmezliđi ejeksiyon fraksiyon deđerine göre iki türe ayırır. İlki, daha önce sol ventriküler (LV) sistolik disfonksiyon veya sistolik kalp yetmezliđine bađlı kalp yetmezliđi olarak bilinen ve %40'tan daha küçük bir ejeksiyon fraksiyonu ile karakterize edilen azalmış ejeksiyon fraksiyonuna bađlı kalp yetmezliđidir. İkinci-si, eskiden diyastolik kalp yetmezliđi veya normal ejeksiyon fraksiyonlu kalp yetmezliđi olarak adlandırılan, korunmuş ejeksiyon fraksiyonu ile kalp yetmezliđidir. Bu durumda, sol ventrikül sistol sırasında normal olarak kasılır, ancak ventrikül serttir ve diyastol sırasında normal olarak gevşememekte, dolayısıyla dolumu bozmaktadır

1 Giriş

Kalp yetmezliđi, kalp duvarındaki kasların solup büyüdüđu ve kalbin kan pompalamasını sınırladıđı durumdur. Kalbin ventrikülleri esnek olmayabilir ve atımlar arasında düzgün bir şekilde dolmayabilir. Zamanla kalp, vücuttaki kanın uygun talebini karşılayamaz hale gelir ve bunun sonucunda kiři nefes almakta güçlük çekmeye başlar.

*20080699, [Github Repo](#)

Kalp yetmezliđinin ana nedeni koroner kalp hastalıđı , diyabet , yüksek tansiyon ve HIV, alkol kötüye kullanımı veya kokain, tiroid bozuklukları, vücutta E vitamini fazlalığı, radyasyon veya kemoterapi vb . Kardiyovasküler Kalp Hastalıđı (KKH) řu anda dünya çapında ölümlerin %31'ine neden olan en önemli nedendir. Pakistan da KKH prevalansının önemli ölçüde arttığı ölkeler listesine dahil edilmiştir.

Bu bölgede kalp yetmezliğine odaklanan çalışmaların görece azlığına ek olarak, Pakistan'daki diyet kalıpları Hindistan, Bangladeř, Nepal ve Sri Lanka gibi diđer Güney Asya ölkelerinden farklı olduğundan, bu çalışma Pakistan bağlamında özel bir öneme sahiptir.

1.1 Çalışmanın Amacı

Bu çalışmada amaç bir kalp yetmezliği hastasının hayatta kalma şansını etkileyen faktörleri incelemek ve vaka örnekleri üzerinden bu hastalığın tedavisinde dikkat edilmesi gereken noktaları açıklamaktır. Ölüm olayını; hastanın kardiyak fonksiyonları, semptomları ve ilaç kullanımı gibi faktörleri ele alarak aralarındaki ilişkiyi matematiksel bir eşitlik ile açıklamaktır. Amaç, kalp yetmezliği hastalarının tedavi edilmesindeki zorlukları anlamak ve tedavi sürecinde kaliteli bir yaşam sürdürebilmeleri için neler yapılması gerektiğini belirtmektir. Ahmad vd. (2017)

Bu çalışmanın temel amacı, çalışma alanı olarak Faysalabad'ı (Pakistan'ın en kalabalık üçüncü şehri) seçerek kalp yetmezliğine bağlı ölüm oranlarını tahmin etmek ve bazı önemli risk faktörleriyle bağlantısını araştırmaktır. Kalbin vücuda yeteri kadar kan pompalamaması sonucu görülen kalp yetmezliği sigara kullanımı, yüksek tansiyon, diyabet, stres gibi faktörler ile daha hızlı ilerlemektedir. Bu hastalığa bağlı hayatta kalma analizi için bir model geliştirmek üzere hastaların belirli bir süre takip edilmesi ve hasta öldüğünde hasta ile ilgili verilerin kayıt altına alınması gerekmektedir.

1.2 Literatür

Kalp yetmezliği, kalp duvarındaki kasların solup büyüdüğü ve kalbin kan pompalamasını sınırladığı durumdur. Kalbin ventrikülleri esnek olmayabilir ve atımlar arasında düzgün bir şekilde dolmayabilir. Zamanla kalp, vücuttaki kanın uygun talebini karşılayamaz hale gelir ve bunun sonucunda kişi nefes almakta güçlük çekmeye başlar.

Kalp yetmezliđinin ana nedeni koroner kalp hastalıđı , diyabet , yüksek tansiyon ve HIV, alkol kötüye kullanımı veya kokain, tiroid bozuklukları, vücutta E vitamini fazlalığı, radyasyon veya kemoterapi vb . Kardiyovasküler Kalp Hastalıđı (KKH) řu anda dünya çapında ölümlerin %31'ine neden olan en önemli nedendir. Pakistan da KKH prevalansının önemli ölçüde arttığı ölkeler listesine dahil edilmiştir. Al-Shifa hastanesinin raporuna göre, 45 yaş üstü Pakistan nüfusunun %33'ünde hipertansiyon var, 45 yaş üstü hastaların %25'inde diyabet var ve Pakistan'da KKH ölümleri yılda yaklaşık 200.000'e ulařtı, yani nüfusun 410/100.000'i). Bütün bunlar kalp yetmezliği prevalansının artmasına neden olur. Pakistan'da kalp yetmezliği hastalarının oranının milyonda 110 olduğu tahmin edilmektedir. Modern çağda ekonomik

ve sosyal sorunların artan stresi, az egzersizle yağlı yiyecekler Pakistan’da kalp yetmezliği prevalansının artmasına neden oluyor.

Bu endişe verici duruma rağmen, Pillai ve Ganapathi bu bölgede kötü ve yağlı beslenme, egzersiz eksikliği ve kötü sağlık politikaları nedeniyle gerekliyen, bu bölgede kalp yetmezliği insidansı ve prevalansına ilişkin güvenilir tahminlerin bulunmadığını bildirdiler. bölge. Sadece batı ülkelerinden elde edilen yaygınlık verilerine dayanan bazı tahminler vardır.

Bu bölgede kalp yetmezliğine odaklanan çalışmaların görece azlığına ek olarak, Pakistan’daki diyet kılıpları Hindistan, Bangladeş, Nepal ve Sri Lanka gibi diğer Güney Asya ülkelerinden farklı olduğundan, bu çalışma Pakistan bağlamında özel bir öneme sahiptir.@ahmad2017survival

Bu çalışmada, Ahmad ve arkadaşları tarafından yayınlanan, kalp yetmezliği olan hastaların tıbbi kayıtlarını içeren bir veri kümesini makine öğrenme algoritmaları ile analiz ederek, hastaların kalp yetmezliğine bağlı olarak hayatta kalma veya ölme durumu tahmini üzerine çalışılmıştır. Bu amaçla, ilgili veri kümesinde bulunan 12 özniteliğin bir kural (One Rule-1Rule), Rastgele Orman (Random Forest RF), Destek Vektör Makinesi (Support Vector Machine SVM), Çok Katmanlı Perseptron (Multi Layer Perceptron MLP) ve Naif Bayes (Navie Bayes) gibi makine öğrenimi yöntemlerini ayrı ayrı beslediği bir mortalite tahmin uygulaması tasarlanmıştır. Daha sonra doğru tahmin oranını arttırmak/çalışma maliyetini azaltmak için korelasyon tabanlı öznitelik seçim algoritması kullanılarak kardiyovasküler hastalıklar için görece anlamlı öznitelikleri ortaya çıkarmak adına öznitelikler seçilmiştir. Erdaş ve Ölçer (t.y.)

2 Veri

2015 yılında toplanan Faysalabad Kardiyoloji Enstitüsü ve Faisalabad’daki (Pencap, Pakistan) Allied Hastanesinde toplanan 299 kalp yetmezliği hastasının tıbbi kayıtlarını içeren bir veri seti kullanılmıştır. Hastaların 105’i kadın, 194’ü erkek ve yaşları 40-95 arasında değişmektedir.

Veri R'a yüklenmiştir.

Verideki değişkenlerin isimlerini değiştirelim

Betimsel istatistiklerini oluşturalım.

##	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet
##	Min. :40.00	Min. :0.0000	Min. : 23.0	Min. :0.0000
##	1st Qu.:51.00	1st Qu.:0.0000	1st Qu.: 116.5	1st Qu.:0.0000
##	Median :60.00	Median :0.0000	Median : 250.0	Median :0.0000
##	Mean :60.83	Mean :0.4314	Mean : 581.8	Mean :0.4181
##	3rd Qu.:70.00	3rd Qu.:1.0000	3rd Qu.: 582.0	3rd Qu.:1.0000
##	Max. :95.00	Max. :1.0000	Max. :7861.0	Max. :1.0000
##	Enjeksiyon	Fraksiyonu	Yüksek Kan Basıncı	Trombositler
##	Min. :14.00	Min. :0.0000	Min. : 25100	Min. :0.500
				Serum Kreatinin

## 1st Qu.:30.00	1st Qu.:0.0000	1st Qu.:212500	1st Qu.:0.900
## Median :38.00	Median :0.0000	Median :262000	Median :1.100
## Mean :38.08	Mean :0.3512	Mean :263358	Mean :1.394
## 3rd Qu.:45.00	3rd Qu.:1.0000	3rd Qu.:303500	3rd Qu.:1.400
## Max. :80.00	Max. :1.0000	Max. :850000	Max. :9.400
## Serum Sodyum	Cinsiyet	Sigara içmek	Zaman
## Min. :113.0	Min. :0.0000	Min. :0.0000	Min. : 4.0
## 1st Qu.:134.0	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.: 73.0
## Median :137.0	Median :1.0000	Median :0.0000	Median :115.0
## Mean :136.6	Mean :0.6488	Mean :0.3211	Mean :130.3
## 3rd Qu.:140.0	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:203.0
## Max. :148.0	Max. :1.0000	Max. :1.0000	Max. :285.0
## Ölüm Olayı			
## Min. :0.0000			
## 1st Qu.:0.0000			
## Median :0.0000			
## Mean :0.3211			
## 3rd Qu.:1.0000			
## Max. :1.0000			

Değişkenlerin standart sapmalarını hesaplıyalım.

[1] 11.89481

[1] 0.4961073

[1] 970.2879

[1] 0.4940671

[1] 11.83484

[1] 0.4781364

[1] 97804.24

[1] 1.03451

[1] 4.412477

[1] 0.4781364

[1] 0.4676704

[1] 77.61421

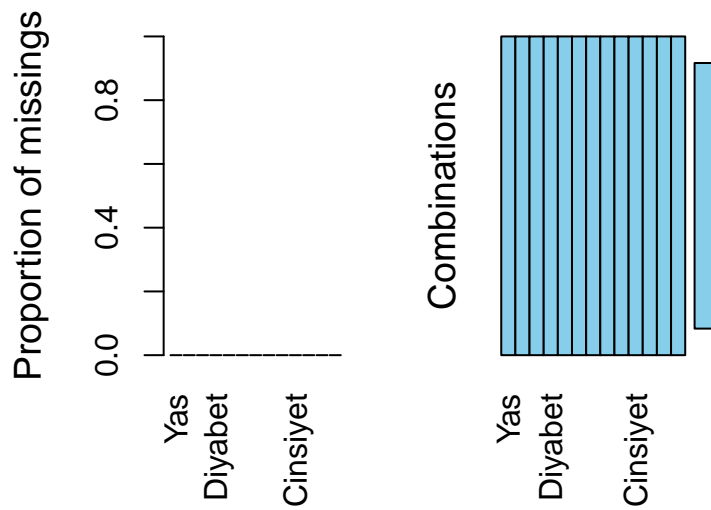
[1] 0.4676704

## [1] "Yaş"	"Anemi"	"Kreatinin Fosfokinaz"
## [4] "Diyabet"	"Enjeksiyon Fraksiyonu"	"Yüksek Kan Basıncı"
## [7] "Trombositler"	"Serum Kreatinin"	"Serum Sodyum"
## [10] "Cinsiyet"	"Sigara içmek"	"Zaman"
## [13] "Ölüm Olayı"		

##	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet	Enjeksiyon Fraksiyonu
## 1	75	0	582	0	20
## 2	55	0	7861	0	38
## 3	65	0	146	0	20
## 4	50	1	111	0	20
## 5	65	1	160	1	20
## 6	90	1	47	0	40

##	Yüksek Kan Basıncı	Trombositler	Serum Kreatinin	Serum Sodyum	Cinsiyet
## 1	1	265000	1.9	130	1
## 2	0	263358	1.1	136	1
## 3	0	162000	1.3	129	1
## 4	0	210000	1.9	137	1
## 5	0	327000	2.7	116	0
## 6	1	204000	2.1	132	1

##	Sigara içmek	Zaman	Ölüm Olayı
## 1	0	4	1
## 2	0	6	1
## 3	1	7	1
## 4	0	7	1
## 5	0	8	1
## 6	1	8	1



```
##
## Missings in variables:
## [1] Variable Count
## <0 rows> (or 0-length row.names)
```

```
##
## Missings per variable:
##           Variable Count
##           Yaş          0
##           Anemi        0
##   Kreatinin Fosfokinaz  0
##           Diyabet      0
## Enjeksiyon Fraksiyonu  0
##   Yüksek Kan Basıncı   0
##           Trombositler  0
##   Serum Kreatinin      0
##           Serum Sodyum  0
##           Cinsiyet     0
##           Sigara içmek  0
##           Zaman        0
##           Ölüm Olayı    0
##
## Missings in combinations of variables:
##           Combinations Count Percent
## 0:0:0:0:0:0:0:0:0:0:0:0:0  299    100
```

```
## [1] "Yaş" "Anemi" "Kreatinin Fosfokinaz"
## [4] "Diyabet" "Enjeksiyon Fraksiyonu" "Yüksek Kan Basıncı"
```

```
## [7] "Trombositler"          "Serum Kreatinin"      "Serum Sodyum"
## [10] "Cinsiyet"              "Sigara içmek"         "Zaman"
## [13] "Ölüm Olayı"
```

```
## Yaş Anemi Kreatinin Fosfokinaz Diyabet Enjeksiyon Fraksiyonu
## 1 75 0 582 0 20
## 2 55 0 7861 0 38
## 3 65 0 146 0 20
## 4 50 1 111 0 20
## 5 65 1 160 1 20
## 6 90 1 47 0 40
## Yüksek Kan Basıncı Trombositler Serum Kreatinin Serum Sodyum Cinsiyet
## 1 1 265000 1.9 130 1
## 2 0 263358 1.1 136 1
## 3 0 162000 1.3 129 1
## 4 0 210000 1.9 137 1
## 5 0 327000 2.7 116 0
## 6 1 204000 2.1 132 1
## Sigara içmek Zaman Ölüm Olayı
## 1 0 4 1
## 2 0 6 1
## 3 1 7 1
## 4 0 7 1
## 5 0 8 1
## 6 1 8 1
```

```
## Yaş Anemi Kreatinin Fosfokinaz Diyabet
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:299 FALSE:299 FALSE:299 FALSE:299
## Enjeksiyon Fraksiyonu Yüksek Kan Basıncı Trombositler Serum Kreatinin
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:299 FALSE:299 FALSE:299 FALSE:299
## Serum Sodyum Cinsiyet Sigara içmek Zaman
## Mode :logical Mode :logical Mode :logical Mode :logical
## FALSE:299 FALSE:299 FALSE:299 FALSE:299
## Ölüm Olayı
## Mode :logical
## FALSE:299
```

3 Yöntem ve Veri Analizi

Bir değişkeni çok fazla sayıda değişken bir araya gelerek o değişkeni etkileyebilmektedirler. Bu değişkenler aynı zamanda kendi aralarında da birbirlerini etkileyebilmektedir. Birden

fazla bağımsız değişken kullanılarak yapılan regresyon analizine “çoklu regresyon analizi (multiple regression analysis)” adı verilmektedir. Çoklu doğrusal regresyon denklemi aşağıda verilmiştir.

$$Y_t = \beta_0 + \beta_N N_t + \beta_P P_t + \beta_I I_t + \varepsilon_t$$

Çoklu Doğrusal Regresyonda değişkenler nicel olmalıdır. Doğrusal Regresyon için varsayımlar vardır. Bu varsayımlardan biri normal dağılım ve korelasyon analizi (doğrusal ilişki olmalıdır) önem taşır.

Bu çalışmada çoklu doğrusal regresyon analizinden yardım alınmıştır. Çoklu Doğrusal Regresyonun amacı, amacımız olan ölüm olayı ile diğer faktörler arasındaki en iyi denklemi çıkarabilmektir.

Normallik testi

Pek çok çalışmada normallik varsayımı kontrol edildikten sonra yapılacak analizlerin çeşitlerine karar verilir. Veri setinin normal dağılımı durumunda dağılımın şekli simetrik bir çan eğrisine benzemektedir.

Aşağıdaki verilerin normallik testinde çıkan değerleri verilmektedir. H0: Veriler Normal Dağılıma Sahiptir. H1: Veriler Normal Dağılıma Sahip Değildir.

```
##
##  Shapiro-Wilk normality test
##
## data:  veri$Yaş
## W = 0.97547, p-value = 5.35e-05

##
##  Shapiro-Wilk normality test
##
## data:  veri$Anemi
## W = 0.62961, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Kreatinin Fosfokinaz`
## W = 0.51426, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  veri$Diyabet
## W = 0.62665, p-value < 2.2e-16
```



```

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Enjeksiyon Fraksiyonu`
## W = 0.94732, p-value = 7.216e-09

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Yüksek Kan Basıncı`
## W = 0.60343, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  veri$Trombositler
## W = 0.91151, p-value = 2.883e-12

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Serum Kreatinin`
## W = 0.55147, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Serum Sodyum`
## W = 0.93903, p-value = 9.215e-10

##
##  Shapiro-Wilk normality test
##
## data:  veri$Cinsiyet
## W = 0.60343, p-value < 2.2e-16

##
##  Shapiro-Wilk normality test
##
## data:  veri$`Sigara içmek`
## W = 0.58814, p-value < 2.2e-16

```

```
##
## Shapiro-Wilk normality test
##
## data: veri$Zaman
## W = 0.94678, p-value = 6.285e-09
```

```
##
## Shapiro-Wilk normality test
##
## data: veri$`Ölüm Olayı`
## W = 0.58814, p-value < 2.2e-16
```

Yukarıdaki verilen değerlere bakıldığında tüm değerler 0.05'ten küçüktür (p-value < 0.05). Bu değerler 0.05'ten küçük olduğundan H0 red edilemez. Yani Veriler Normal Dağılıma Sahiptir.

Korelasyon Matrisi oluşturalım

Korelasyon analizi, iki nicel değişken arasındaki doğrusal ilişkinin gücü ve yönü hakkında sayısal bilgi verir.

Aşağıdaki verilerin korelasyon testinde çıkan değerleri vermektedir.

##	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet
## Yaş	1.00000000	0.08800644	-0.081583900	-0.101012385
## Anemi	0.08800644	1.00000000	-0.190741030	-0.012729046
## Kreatinin Fosfokinaz	-0.08158390	-0.19074103	1.000000000	-0.009638514
## Diyabet	-0.10101239	-0.01272905	-0.009638514	1.000000000
## Enjeksiyon Fraksiyonu	0.06009836	0.03155697	-0.044079554	-0.004850310
## Yüksek Kan Basıncı	0.09328868	0.03818200	-0.070589980	-0.012732382
## Trombositler	-0.05235437	-0.04378555	0.024463389	0.092192828
## Serum Kreatinin	0.15918713	0.05217360	-0.016408480	-0.046975315
## Serum Sodyum	-0.04596584	0.04188161	0.059550156	-0.089550619
## Cinsiyet	0.06542952	-0.09476896	0.079790629	-0.157729504
## Sigara içmek	0.01866787	-0.10728984	0.002421235	-0.147173413
## Zaman	-0.22406842	-0.14141398	-0.009345653	0.033725509
## Ölüm Olayı	0.25372854	0.06627010	0.062728160	-0.001942883
##	Enjeksiyon Fraksiyonu	Yüksek Kan Basıncı	Trombositler	
## Yaş	0.06009836	0.093288685	-0.05235437	
## Anemi	0.03155697	0.038182003	-0.04378555	
## Kreatinin Fosfokinaz	-0.04407955	-0.070589980	0.02446339	
## Diyabet	-0.00485031	-0.012732382	0.09219283	
## Enjeksiyon Fraksiyonu	1.00000000	0.024444731	0.07217747	
## Yüksek Kan Basıncı	0.02444473	1.000000000	0.04996348	
## Trombositler	0.07217747	0.049963481	1.00000000	
## Serum Kreatinin	-0.01130247	-0.004934525	-0.04119808	

## Serum Sodyum	0.17590228	0.037109470	0.06212462
## Cinsiyet	-0.14838597	-0.104614629	-0.12512048
## Sigara içmek	-0.06731457	-0.055711369	0.02823445
## Zaman	0.04172924	-0.196439479	0.01051391
## Ölüm Olayı	-0.26860331	0.079351058	-0.04913887
##	Serum Kreatinin	Serum Sodyum	Cinsiyet Sigara içmek
## Yaş	0.159187133	-0.045965841	0.065429524 0.018667868
## Anemi	0.052173604	0.041881610	-0.094768961 -0.107289838
## Kreatinin Fosfokinaz	-0.016408480	0.059550156	0.079790629 0.002421235
## Diyabet	-0.046975315	-0.089550619	-0.157729504 -0.147173413
## Enjeksiyon Fraksiyonu	-0.011302475	0.175902282	-0.148385965 -0.067314567
## Yüksek Kan Basıncı	-0.004934525	0.037109470	-0.104614629 -0.055711369
## Trombositler	-0.041198077	0.062124619	-0.125120483 0.028234448
## Serum Kreatinin	1.000000000	-0.189095210	0.006969778 -0.027414135
## Serum Sodyum	-0.189095210	1.000000000	-0.027566123 0.004813195
## Cinsiyet	0.006969778	-0.027566123	1.000000000 0.445891712
## Sigara içmek	-0.027414135	0.004813195	0.445891712 1.000000000
## Zaman	-0.149315418	0.087640000	-0.015608220 -0.022838942
## Ölüm Olayı	0.294277561	-0.195203596	-0.004316376 -0.012623153
##	Zaman	Ölüm Olayı	
## Yaş	-0.224068420	0.253728543	
## Anemi	-0.141413982	0.066270098	
## Kreatinin Fosfokinaz	-0.009345653	0.062728160	
## Diyabet	0.033725509	-0.001942883	
## Enjeksiyon Fraksiyonu	0.041729235	-0.268603312	
## Yüksek Kan Basıncı	-0.196439479	0.079351058	
## Trombositler	0.010513909	-0.049138868	
## Serum Kreatinin	-0.149315418	0.294277561	
## Serum Sodyum	0.087640000	-0.195203596	
## Cinsiyet	-0.015608220	-0.004316376	
## Sigara içmek	-0.022838942	-0.012623153	
## Zaman	1.000000000	-0.526963779	
## Ölüm Olayı	-0.526963779	1.000000000	

Korelasyon değeri eğer pozitif yönlü bir değer alırsa pozitif yönlü bir ilişki, negatif bir değer alır ise her iki değişken arasında negatif yönlü bir ilişki vardır. Tablo3. daki değerler faktörlerin birbiriyle aralarındaki ilişkinin korelasyon değerleri ve aldığı işarete göre de pozitif negatif yönünü belirtmektedir.

Örneğin; Yaş ile Diyabet arasındaki korelasyon katsayısı -0.101012385 0.10 olarak hesaplanmıştır. Yaş ile Diyabet arasında negatif düşük yönlü bir ilişki vardır.

Çoklu Doğrusal Regresyon

R programlamada gerekli kodlar yazılarak sonuçlar görüntülenmiştir. Bu sonuçlar tablo halinde getirilmiştir.Çoklu doğrusal regresyon analizinin sonuçları verilmiştir.

3.1 R işlevi lm(), doğrusal modelin beta katsayılarını belirlemek için kullanılabilir:

```
##
## Call:
## lm(formula = veri$`Ölüm Olayı` ~ ., data = veri)
##
## Coefficients:
##              (Intercept)                Yaş                Anemi
##              1.664e+00              5.767e-03              -2.766e-03
## `Kreatinin Fosfokinaz`              Diyabet `Enjeksiyon Fraksiyonu`
##              3.427e-05              1.928e-02              -9.834e-03
## `Yüksek Kan Basıncı`              Trombositler `Serum Kreatinin`
##              -1.430e-02              -8.370e-08              8.527e-02
## `Serum Sodyum`              Cinsiyet `Sigara içmek`
##              -7.599e-03              -6.369e-02              -5.733e-03
##              Zaman
##              -2.733e-03

##
## Call:
## lm(formula = veri$`Ölüm Olayı` ~ ., data = veri)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80866 -0.28041 -0.04205  0.24742  0.96983
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.664e+00  6.954e-01   2.392  0.01738 *
## Yaş              5.767e-03  1.867e-03   3.088  0.00221 **
## Anemi            -2.766e-03  4.438e-02  -0.062  0.95035
## `Kreatinin Fosfokinaz` 3.427e-05  2.247e-05   1.525  0.12840
## Diyabet          1.928e-02  4.410e-02   0.437  0.66236
## `Enjeksiyon Fraksiyonu` -9.834e-03  1.844e-03  -5.333 1.96e-07 ***
## `Yüksek Kan Basıncı`  -1.430e-02  4.565e-02  -0.313  0.75438
## Trombositler      -8.370e-08  2.208e-07  -0.379  0.70492
## `Serum Kreatinin`     8.527e-02  2.123e-02   4.017 7.54e-05 ***
## `Serum Sodyum`       -7.599e-03  5.024e-03  -1.513  0.13149
## Cinsiyet          -6.369e-02  5.108e-02  -1.247  0.21353
## `Sigara içmek`       -5.733e-03  5.119e-02  -0.112  0.91091
## Zaman             -2.733e-03  2.903e-04  -9.415 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.3646 on 286 degrees of freedom
## Multiple R-squared:  0.4168, Adjusted R-squared:  0.3924
## F-statistic: 17.04 on 12 and 286 DF,  p-value: < 2.2e-16
```

Residual standard error: 0.3646 on 286 degrees of freedom Multiple R-squared: 0.4168, Adjusted R-squared: 0.3924 F-statistic: 17.04 on 12 and 286 DF, p-value: < 2.2e-16

Modelin anlamlı olup olmadığını çıkan sonuçlara bakarak yorum yapmaya çalışalım;

H0: Model anlamlı değildir. H1: Model anlamlıdır.

Hesaplanan p değerine bakıldığında p-value değeri < 2.2e-16 olarak hesaplanmıştır. Bu değer 0.01'den bile küçük bir değer olduğunu söyleyebiliriz. P-value değeri 0.05'ten küçük olduğundan (p-value < 0.05) H0 hipotezi red edilir. Yani modelimizin anlamlı olduğunu söyleyebiliriz.

Modelin anlamlı olup olmadığını söyleyebilmek için F istatistik değerine bakarak da karar verebiliriz.

$F_{((0.05;12;286))} = 3.86$ 'dır. (F tablosuna bakıldı)

$F_{\text{Hesap}} = 17.04$ yani $F_{\text{Hesap}} > F_{((0.05;12;286))}$ olduğundan H0 hipotezimiz reddedilir. Yani buradan da modelimizin anlamlı olduğunu söyleyebiliriz.

Katsayıların anlamlılıklarını test etmezsek doğru bir analiz yapmış olmayız bu yüzden anlamlılıklarını test edelim:

H0: Sabit terim anlamlı değildir. H1: Sabit terim anlamlıdır.

$t_{\text{Hesap}} = 2.392$

$t_{((0.025;286))} = 1.960$ (t tablosuna bakıldı)

$t_{\text{Hesap}} > t_{((0.025;286))}$ olduğundan H0 hipotezimiz reddedilir yani Sabit terimin anlamlı olduğunu söyleyebiliriz.

Regresyon modeli;

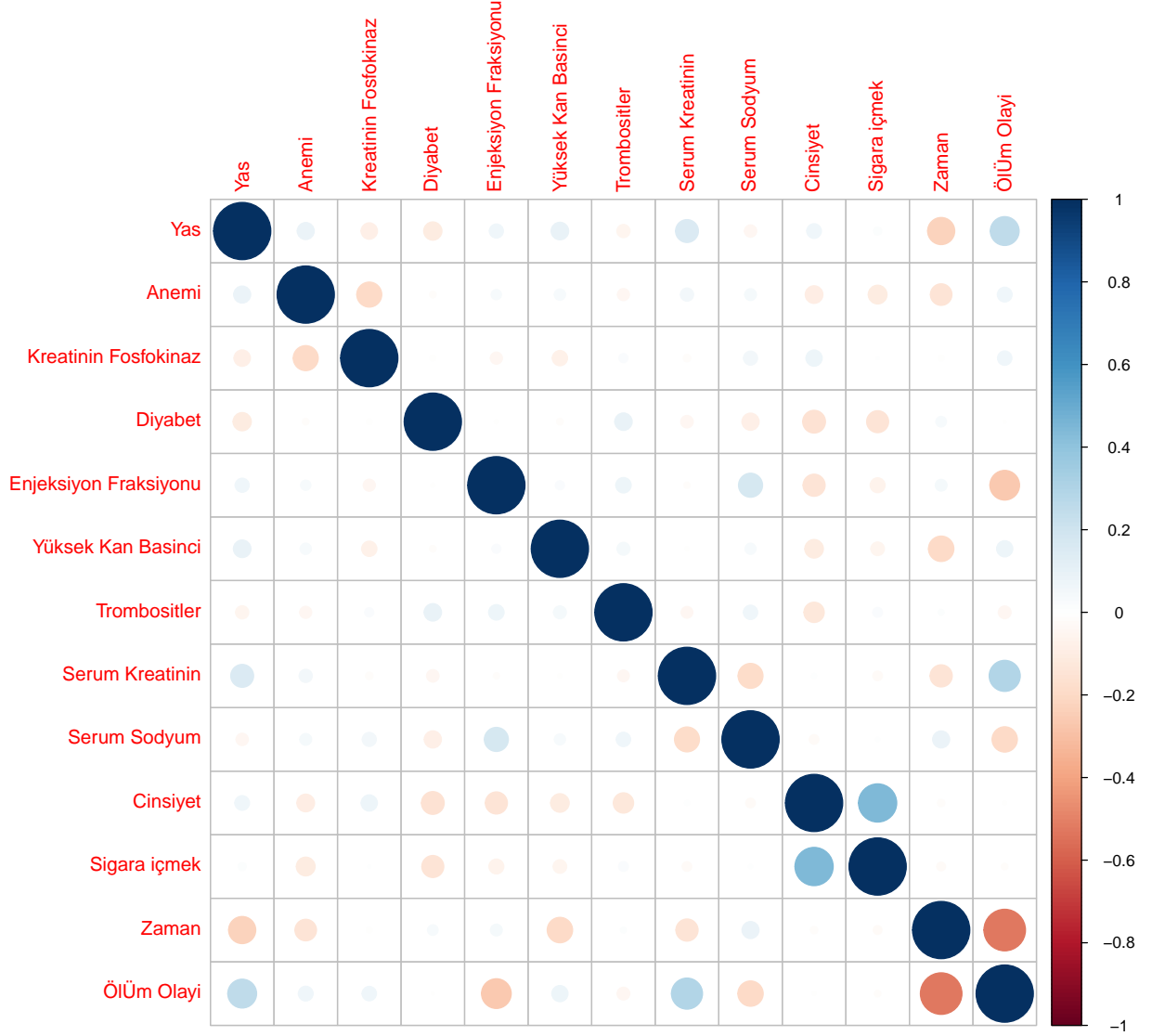
$$Y = (1.664e+00) + (5.767e-03) X_1 + (-9.834e-03) X_2 + (8.527e-02) X_3 + (-2.733e-03) X_4 +$$
 şeklinde model oluşturulabilir.

R2 yani çoklu belirtme katsayısı, bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişkenle açıklanabileceğini gösterir. Bu değer 0 ile 1 arasında değer alır ve bu değer 1'e yakın olması bağımsız değişkenin bağımlı değişkeni iyi bir şekilde açıkladığı anlamına gelir ve bu da bizim istediğimiz bir durumdur. Bu bilgileri verdikten sonra çözümlememize dönecek olursak bizim çoklu belirtme katsayımız 0.4168 olarak açıklanmıştır.

Tüm değişkenlerin birbirine bağlılık derecesini daha net bir şekilde görülebilir. Grafik üzerindeki koyu mavi noktalar güçlü pozitif korelasyonları, koyu kırmızı noktalar ise güçlü negatif korelasyonları göstermektedir. Diğer daireler de renklerine ve büyüklüklerine göre korelasyonun gücünü ve yönünü belirtmektedir.

Örneğin; Ölüm Olayı ile zaman arasında negatif yüksek korelasyonlu bir ilişki vardır.
Korelasyon Grafiği

Çeşitli grafik yöntemleriyle de yapılan analizleri destekleyici ve göze hitabı arttırmayı sağlayalım.



##	Yaş	Anemi	Kreatinin Fosfokinaz	Diyabet	Enjeksiyon Fraksiyonu
## 1	75	0		582	0
## 2	55	0		7861	0
## 3	65	0		146	0
## 4	50	1		111	0
## 5	65	1		160	1
## 6	90	1		47	0
##	Yüksek Kan Basıncı	Trombositler	Serum Kreatinin	Serum Sodyum	Cinsiyet
## 1		1	265000	1.9	130

```

## 2      0      263358      1.1      136      1
## 3      0      162000      1.3      129      1
## 4      0      210000      1.9      137      1
## 5      0      327000      2.7      116      0
## 6      1      204000      2.1      132      1
##   Sigara içmek Zaman Ölüm Olayı
## 1      0      4      1
## 2      0      6      1
## 3      1      7      1
## 4      0      7      1
## 5      0      8      1
## 6      1      8      1

## 'data.frame':   299 obs. of  13 variables:
##  $ Yaş          : num  75 55 65 50 65 90 75 60 65 80 ...
##  $ Anemi         : int  0 0 0 1 1 1 1 1 0 1 ...
##  $ Kreatinin Fosfokinaz : int  582 7861 146 111 160 47 246 315 157 123 ...
##  $ Diyabet       : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ Enjeksiyon Fraksiyonu: int  20 38 20 20 20 40 15 60 65 35 ...
##  $ Yüksek Kan Basıncı  : int  1 0 0 0 0 1 0 0 0 1 ...
##  $ Trombositler       : num  265000 263358 162000 210000 327000 ...
##  $ Serum Kreatinin     : num  1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4 ...
##  $ Serum Sodyum        : int  130 136 129 137 116 132 137 131 138 133 ...
##  $ Cinsiyet           : int  1 1 1 1 0 1 1 1 0 1 ...
##  $ Sigara içmek       : int  0 0 1 0 0 1 0 1 0 1 ...
##  $ Zaman              : int  4 6 7 7 8 8 10 10 10 10 ...
##  $ Ölüm Olayı        : int  1 1 1 1 1 1 1 1 1 1 ...

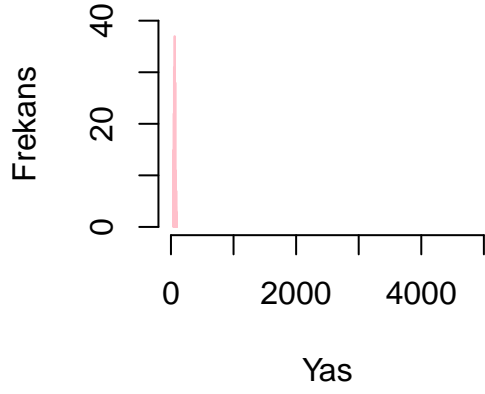
```

Şekile bakıldığında tüm değişkenlerin birbirine bağlılık derecesini daha net bir şekilde görülebilir. Grafik üzerindeki koyu mavi noktalar güçlü pozitif korelasyonları, koyu kırmızı noktalar ise güçlü negatif korelasyonları göstermektedir. Diğer daireler de renklerine ve büyüklüklerine göre korelasyonun gücünü ve yönünü belirtmektedir.

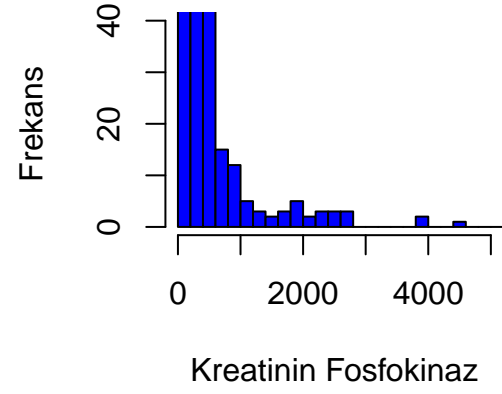
Örneğin; Ölüm Olayı ile zaman arasında negatif yüksek korelasyonlu bir ilişki vardır.

Histogram grafikleri

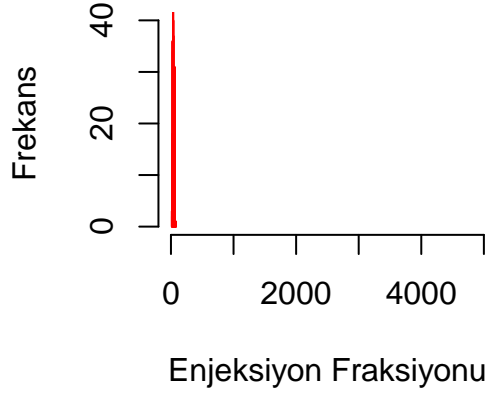
weightlbs icin Histogram



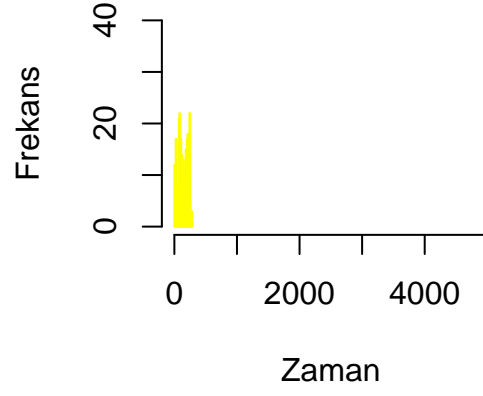
weightlbs icin Histogram



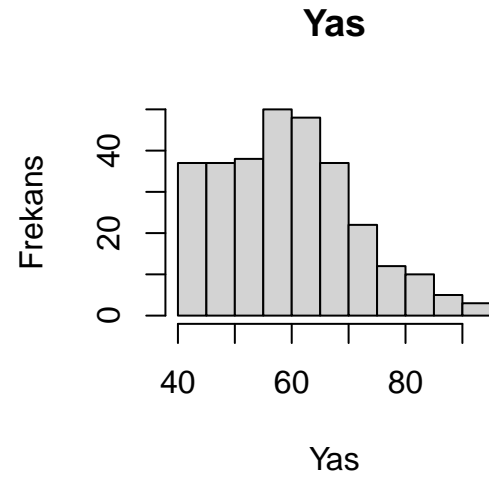
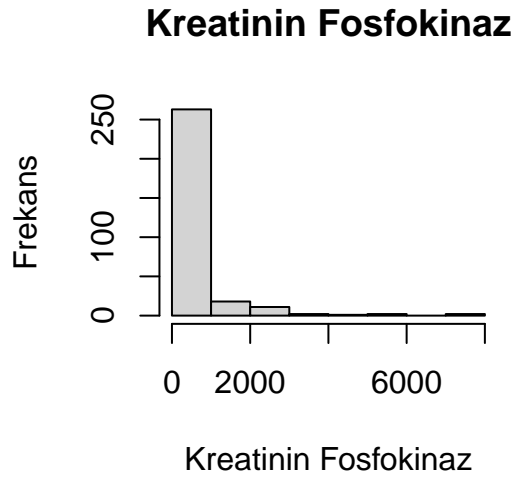
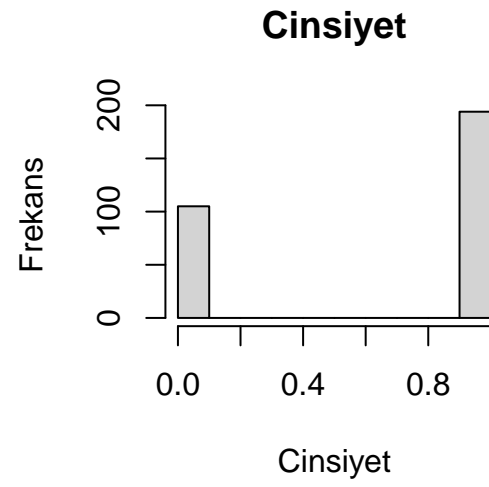
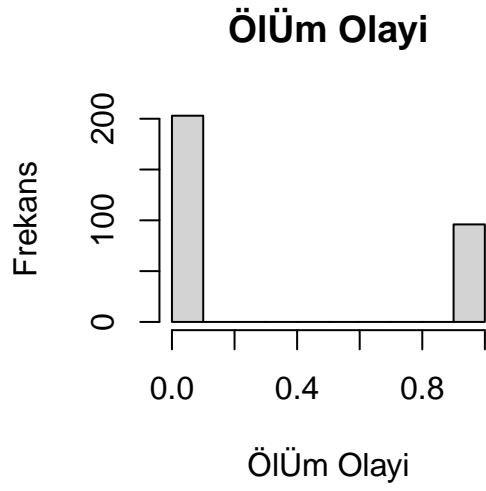
weightlbs icin Histogram



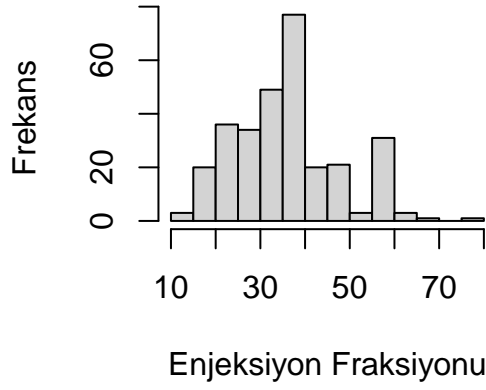
weightlbs icin Histogram



R programında farklı histogram kodu kullanılarak çıkartılmıştır.



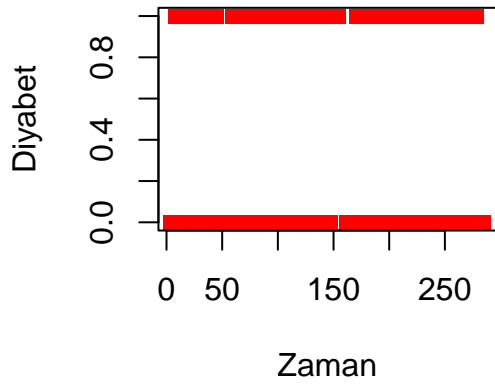
Enjeksiyon Fraksiyonu



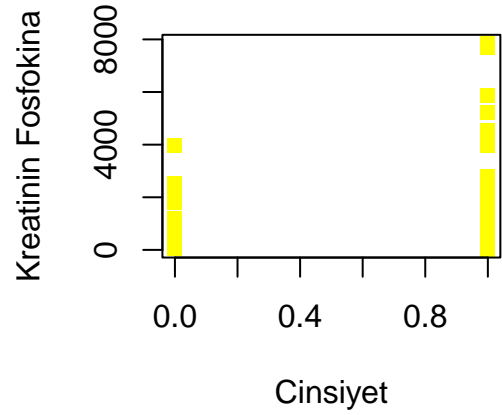
Histogram grafikleri bize ortalama, standart sapma ve örneklem sayısı hakkında bilgi veren bir grafikdir. Verilerin dağılımları hakkında da ön bilgi oluşturabilir.

Dağılım grafikleri

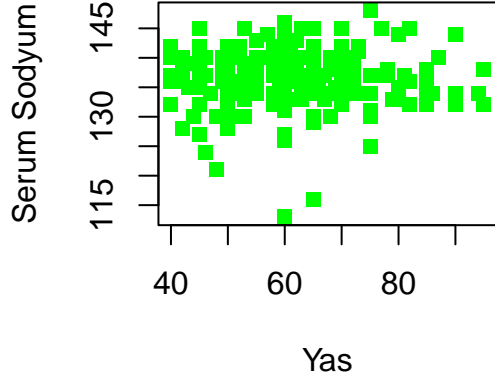
Zaman – Diyabet



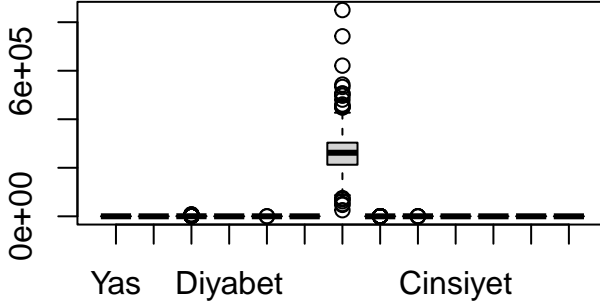
Cinsiyet – Kreatinin Fosfokina



Yas – Serum Sodyum



Kutu grafiği



4 Sonuç

Kalp yetmezliğinin ölüme neden olan sebeplerden biri olduğu bilinmektedir. Yapılan araştırmada istenilen şey kalp yetmezliği hastalarının tedavi edilmesindeki zorlukları anlamak ve tedavi sürecinde kaliteli bir yaşam sürdürebilmeleri için neler yapılması gerektiğini belirlemektir. Bizim araştırdığımız ise ölüm olayı ile faktörler arasındaki ilişkiyi matematiksel denkleme dökmektir. Araştırmada Çoklu Doğrusal Regresyon Analizi kullanılmıştır. İlk önce Normallik testi uygulanarak verilerin normal dağılıp dağılmadığı irdelenmiştir. Sonucunda ise verilerin normal dağılım gösterdiği görülmüştür. Daha sonra korelasyon matrisi oluşturulmuştur. Korelasyon analizinde ise ölüm olayı ile yaş arasında pozitif bir ilişki varken ölüm

olayı ile zaman arasında negatif bir ilişki görülmektedir. Daha sonra Çoklu doğrusal regresyon uygulandığında R^2 değeri 0.4168 olarak hesaplanmıştır. Bu değerin 0.30'dan büyük olması yeterli bir sonuçtur. Sonuç olarak çoklu doğrusal regresyon modelinin ölüm olayını tahminini açıklama adına başarılı bir yöntem olduğu tespit edilmiştir. Araştırmalarımıza göre yaş, anemi, enjeksiyonel fraksiyonu, serum kreatinin ve zaman ölüm olayı için önemli faktörlerdir. Ölüm olayına hangi değişkenlerin etkilediği tespit edebilmiş belirlediğiniz araştırma sorusuna cevap verdiğini ve ileride bu çalışmanın ileri tahmin yöntemleri ile geliştirilmesi önerilmektedir. Potur ve Erginel (2021)

5 Kaynakça

- Ahmad, T., Munir, A., Bhatti, S. H., Aftab, M. ve Raza, M. A. (2017). Survival analysis of heart failure patients: A case study. *PloS one*, 12(7), e0181001.
- Erdaş, Ç. B. ve Ölçer, D. (t.y.). Kalp Yetmezliği Hastalarının Hayatta Kalma Tespiti İçin Makine Öğrenmesi Tabanlı Bir Yaklaşım A Machine Learning-Based Approach to Detect Survival of Heart Failure Patients.
- Potur, E. A. ve Erginel, N. (2021). Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (24), 112-118.