



IST-401 ÇOK DEĞİŞKENLİ İSTATİSTİKSEL YÖNTEMLER PROJE ÖDEVİ

Tarık Kütük-20181101048



18 EKİM 2021

MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ

VERİ YAPISI VE DEĞİŞKENLERİN AÇIKLAMASI:

Bu raporda analiz edilecek veri, 303 gözlem ve 12 değişkenden oluşmaktadır. Verinin kaynağı “kaynakça” bölümünde belirtilmiş olup; Budapeşte’deki Macar Kardiyoloji Enstitüsü, İsveç’teki Üniversite Hastanesi, Basel’deki Üniversite Hastanesi ve Long Beach ve Cleveland’daki V.A. Tıp Merkezindeki belirli kardiyovasküler olayları tahmin etmek ve kalp sağlığına ilişkin net göstergeler bulabilmek için oluşturulmuş bir veridir. Veride kayıp gözlem yoktur.

Ocak 2021’de yayınlanmış verinin değişkenleri şöyledir:

- a. yas: Gözlemlenen hastaların yaşı
- b. cinsiyet: Gözlemlenen hastaların cinsiyeti
 - b.1. 0: Kadın
 - b.2. 1: Erkek
- c. drgn_kbsnc: Hasta dinlenirken/kan akışı durgunken kan basıncı
- d. klstrl: Hastanın kolesterolü
- e. mx_klp_ats: Hastanın ölçülen maksimum kalp atış hızı
- f. oldpeak: Kan akışı durgunken egzersizin neden olduğu ST depresyonu
- g. kagt: Kalp Ağrısı Türü
 - g.1. 0 Tipik Angina
 - g.2. 1 Atipik Angina
 - g.3. 2 Anginal Olmayan Ağrı
 - g.4. 3 Asemptomatik Ağrı
- h. ac_kbsnc: Hasta Açken Kan Basıncı
 - h.1. 0: 120mg/dl’den az
 - h.2. 1: 120mg/dl’den fazla
- i. restecg: Hastanın Elektrokardiyografik Sonuçları
 - i.1. 0: Normal
 - i.2. 1: ST-T Dalga Anormalliği var
 - i.3. 2: Estes Kriterlerine göre olası veya kesin sol ventrikül hipertrofi var
- j. exang: Egzersize Bağlı Angina Varlığı
 - j.1. 0: Yok
 - j.2. 1: Var
- k. slope: ST segmentinin eğimi
 - k.1. 0: Yukarı Eğimli
 - k.2. 1: Eğimsiz-Düz
 - k.3. 2: Aşağı Eğimli
- l. target: Kalp Hastalığı Teşhisi

1.1. 0: Sağlıklı

1.2. 1: Hasta

yas	cinsiyet	drgn_kbsnc	klstrl	mx_klp_ats	oldpeak	anginal	kagt
Min. :29.00	kadın: 96	Min. : 94.0	Min. :126.0	Min. : 71.0	Min. : 0.000	anginal :143	
1st Qu.:47.50	erkek:207	1st Qu.:120.0	1st Qu.:211.0	1st Qu.:133.5	1st Qu.: 0.000	non_anginal : 50	
Median :55.00		Median :130.0	Median :240.0	Median :153.0	Median : 4.000	atipik_anginal: 87	
Mean :54.37		Mean :131.6	Mean :246.3	Mean :149.6	Mean : 8.644	asemptomatik : 23	
3rd Qu.:61.00		3rd Qu.:140.0	3rd Qu.:274.5	3rd Qu.:166.0	3rd Qu.:14.000		
Max. :77.00		Max. :200.0	Max. :564.0	Max. :202.0	Max. :62.000		
ac_kbsnc120		restecg	exang	slope	target		
120 mg'dan az :258	normal	:147	yok:204	yukarı : 21	sağlıklı:138		
120 mg'dan fazla: 45	ST-T dalga anormal:152	var: 99	eğimsiz:140	hasta :165			
	hipertrofi : 4		aşağı :142				

Yukarıda değişkenlere ait bazı özetleyici bilgiler verilmiştir.

Yapılacak bütün analizler %95 güven düzeyinde ($\alpha=0.05$) yapılacaktır.

ÇOK DEĞİŞKENLİ VARYANS ANALİZİ

Tek Yönlü MANOVA:

Yaş ve durgun kan basıncı bağımlı, slope bağımsız değişken alınmış ve analize başlanmıştır.

Varsayımlardan ilki olan çok değişkenli normallik için,

H0: Değişkenler çok değişkenli normal dağılmaktadır.

HA: Değişkenler çok değişkenli normal dağılmamaktadır.

```
+ mshapiro_test()
# A tibble: 1 x 2
  statistic p.value
  <dbl>     <dbl>
1      0.978 0.000129
```

R'da yapılan Shapiro-Wilk testinin sig. değerinden H0 reddedilir. Değişkenler çoklu normalliğe uymamaktadır. Değişkenlere ln ve karekök dönüşümleri yapıldıktan sonra da normallik testi yapılmış fakat normal dağılmadıkları görülmüştür. Değişkenlerin dönüşümsüz halleriyle devam edilecek ve analizin devamlılığı için değişkenlerin çoklu normal dağıldığı varsayılacaktır.

İkinci varsayım olan varyansların homojenliği için,

H0: Gruplar arası varyans kovaryans matrisleri eşittir

HA: Gruplar arası varyans kovaryans matrisleri eşit değildir.

```
> box_m(data[, c("yas", "drgn_kbsnc")], data$slope)
# A tibble: 1 x 4
  statistic p.value parameter method
  <dbl>     <dbl>     <dbl> <chr>
1      6.29    0.391         6 Box's M-test for Homogeneity of Covariance Matrices
> |
```

R’da yapılan Box’s M testinin sig. değerinden H0 reddedilemez. Gruplar arası varyanslar homojendir ve analizin bu varsayımı sağlanmıştır.

Between-Subjects Factors

	Value Label	N
slope 0	yukarı eğimli	21
1	eğimsiz-düz	140
2	aşağı eğimli	142

ST segmentinin eğimi değişkenindeki gözlemlerin dağılımı verilmiştir. Burada gözlemlerin kategorilere eşit dağılmadığı dikkat çekmektedir.

Descriptive Statistics

	slope	Mean	Std. Deviation	N
yas	yukarı eğimli	55,38	8,316	21
	eğimsiz-düz	56,10	8,437	140
	aşağı eğimli	52,51	9,490	142
	Total	54,37	9,082	303
drgn_kbsnc	yukarı eğimli	139,48	22,576	21
	eğimsiz-düz	132,10	17,420	140
	aşağı eğimli	129,99	16,591	142
	Total	131,62	17,538	303

ST segmenti eğimi yukarı olan hastaların yaş ortalaması 55, eğimsiz olan hastaların yaş ortalaması 56 ve aşağı eğimli olan hastaların yaş ortalaması ise 53 olduğu görülmektedir. Sırayla standart sapmaları ise 8,32-8,43-9,5 olduğu görülmüştür.

ST segmenti eğimi yukarı olan hastaların durgun kan basıncı ortalamaları 139,48, eğimsiz olan hastaların yaş ortalaması 132,10 ve aşağı olan hastaların yaş ortalaması ise 129,99 olduğu görülmektedir. Sırayla standart sapmaları ise 22,58-17,42-16,59 olduğu görülmüştür.

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,974	5547,437 ^b	2,000	299,000	,000	,974
	Wilks' Lambda	,026	5547,437 ^b	2,000	299,000	,000	,974
	Hotelling's Trace	37,107	5547,437 ^b	2,000	299,000	,000	,974
	Roy's Largest Root	37,107	5547,437 ^b	2,000	299,000	,000	,974
slope	Pillai's Trace	,052	3,968	4,000	600,000	,003	,026
	Wilks' Lambda	,949	3,968 ^b	4,000	598,000	,003	,026
	Hotelling's Trace	,053	3,967	4,000	596,000	,003	,026
	Roy's Largest Root	,040	5,977 ^c	2,000	300,000	,003	,038

a. Design: Intercept + slope

b. Exact statistic

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

H0: $\mu_1=\mu_2=\mu_3$ / Gruplar arasında bağımlı değişkenler açısından anlamlı fark yoktur.

HA: Gruplar arasında bağımlı değişkenler açısından anlamlı fark vardır.

Çok değişkenli testlere baktığımızda, temel varsayımlar sağlanıyorsa Wilk's Lambda, tüm varsayımlar sağlanıyorsa Roy's Largest Root, örneklem küçük veya kategoriler arası gözlem sayısı farklılığı olan durumlarda ise Pilla's Trace testi kullanılır.

Slope değişkenindeki kategorilerin gözlem sayıları arasında farklılık olduğu için Pillai's Trace testinin sig. değerinden H0 reddedilir. Gruplar arası anlamlı bir fark vardır.

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
yas	1,459	2	300	,234
drgn_kbsnc	1,293	2	300	,276

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

a. Design: Intercept + slope

H0: Gruplar arası varyanslar eşittir.

HA: Gruplar arası varyanslar eşit değildir.

Levene's Testi sig değerlerinden iki değişken için de varyansların eşit olduğu görülmüştür.

H0 reddedilemez.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	yas	933,291 ^a	2	466,646	5,839	,003	,037
	drgn_kbsnc	1704,278 ^b	2	852,139	2,803	,062	,018
Intercept	yas	435117,363	1	435117,363	5444,174	,000	,948
	drgn_kbsnc	2609173,643	1	2609173,643	8584,048	,000	,966
slope	yas	933,291	2	466,646	5,839	,003	,037
	drgn_kbsnc	1704,278	2	852,139	2,803	,062	,018
Error	yas	23977,045	300	79,923			
	drgn_kbsnc	91186,831	300	303,956			
Total	yas	920487,000	303				
	drgn_kbsnc	5342310,000	303				
Corrected Total	yas	24910,337	302				
	drgn_kbsnc	92891,109	302				

a. R Squared = ,037 (Adjusted R Squared = ,031)

b. R Squared = ,018 (Adjusted R Squared = ,012)

H0: Yaş açısından kategoriler arasında anlamlı bir fark yoktur.

HA: Yaş açısından kategoriler arasında anlamlı bir fark vardır.

Bu hipotezde yaş değişkeninin sig. değerinin (0,003), alfadan (0,05)den küçük olmasıyla H0 reddedilir. Yaş, ST segmenti eğimi grupları arasında bir farklılığa sebep olmuştur.

H0: Durgun kan basıncı açısından kategoriler arasında anlamlı bir fark yoktur.

HA: Durgun kan basıncı açısından kategoriler arasında anlamlı bir fark vardır.

Durgun kan basıncı sig. değerinin (0,062), alfadan (0,05) büyük olması sebebiyle H0 reddedilemez. Durgun kan basıncı, ST segmenti eğimi grupları arasında bir farklılığa sebep olmamaktadır.

Multiple Comparisons

Dependent Variable		(I) slope	(J) slope	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
yas	Tukey HSD	yukarı eğimli	eğimsiz-düz	-,72	2,092	,937	-5,65	4,21
			aşağı eğimli	2,87	2,090	,355	-2,05	7,80
		eğimsiz-düz	yukarı eğimli	,72	2,092	,937	-4,21	5,65
			aşağı eğimli	3,59*	1,065	,002	1,09	6,10
		aşağı eğimli	yukarı eğimli	-2,87	2,090	,355	-7,80	2,05
			eğimsiz-düz	-3,59*	1,065	,002	-6,10	-1,09
	Tamhane	yukarı eğimli	eğimsiz-düz	-,72	1,950	,977	-5,69	4,25
			aşağı eğimli	2,87	1,982	,403	-2,15	7,90
		eğimsiz-düz	yukarı eğimli	,72	1,950	,977	-4,25	5,69
			aşağı eğimli	3,59*	1,069	,003	1,02	6,16
	Games-Howell	yukarı eğimli	eğimsiz-düz	-,72	1,950	,928	-5,56	4,12
			aşağı eğimli	2,87	1,982	,330	-2,03	7,77
		eğimsiz-düz	yukarı eğimli	,72	1,950	,928	-4,12	5,56
			aşağı eğimli	3,59*	1,069	,003	1,07	6,11
drgn_kbsnc	Tukey HSD	yukarı eğimli	eğimsiz-düz	7,38	4,080	,169	-2,23	16,99
			aşağı eğimli	9,48	4,076	,054	-,12	19,08
		eğimsiz-düz	yukarı eğimli	-7,38	4,080	,169	-16,99	2,23
			aşağı eğimli	2,11	2,076	,568	-2,78	7,00
		aşağı eğimli	yukarı eğimli	-9,48	4,076	,054	-19,08	,12
			eğimsiz-düz	-2,11	2,076	,568	-7,00	2,78
	Tamhane	yukarı eğimli	eğimsiz-düz	7,38	5,142	,417	-5,83	20,58
			aşağı eğimli	9,48	5,119	,213	-3,68	22,65
		eğimsiz-düz	yukarı eğimli	-7,38	5,142	,417	-20,58	5,83
			aşağı eğimli	2,11	2,026	,656	-2,76	6,97
	Games-Howell	yukarı eğimli	eğimsiz-düz	7,38	5,142	,340	-5,47	20,23
			aşağı eğimli	9,48	5,119	,175	-3,33	22,29
		eğimsiz-düz	yukarı eğimli	-7,38	5,142	,340	-20,23	5,47
			aşağı eğimli	2,11	2,026	,552	-2,67	6,88
	Games-Howell	yukarı eğimli	eğimsiz-düz	7,38	5,142	,340	-5,47	20,23
			aşağı eğimli	9,48	5,119	,175	-3,33	22,29
		eğimsiz-düz	yukarı eğimli	-7,38	5,142	,340	-20,23	5,47
			aşağı eğimli	2,11	2,026	,552	-2,67	6,88

Based on observed means.

The error term is Mean Square(Error) = 303,956.

*. The mean difference is significant at the ,05 level.

Çoklu karşılaştırmalar tablosu verilmiştir. Levene's testinde varyansların iki değişken için de eşit olduğu görüldüğü için Tukey testinin sonuçları önemlidir.

Yaşa bakıldığında, ST segmentinin eğimsiz olması ve aşağı eğimli olması arasında anlamlı bir fark olduğu tespit edilmiştir. Hem sig. değerlerinin alfadan küçük olması, hem de güven aralıklarının sıfır değeri içermesiyle bu sonuca varılabilmektedir.

Durgun kan basıncı ve ST segmenti eğimi grupları arasında bir farklılık olmadığı tekrar gözlenir.

Çift Yönlü Manova:

Tek yönlü manova analizindeki bağımlı değişkenler aynı kalırken bağımsız slope değişkeninin yanına bir de kalp ağrısı türü değişkeni eklenerek çift yönlü analiz yapılacaktır.

Varsayımlardan ilki olan çok değişkenli normallik için,

H0: Değişkenler çok değişkenli normal dağılmaktadır.

HA: Değişkenler çok değişkenli normal dağılmamaktadır.

```
+ mshapiro_test()
# A tibble: 1 x 2
  statistic p.value
  <dbl>     <dbl>
1      0.978 0.000129
```

R’da yapılan Shapiro-Wilk testinin sig. değerinden H0 reddedilir. Değişkenler çoklu normalliğe uymamaktadır. Değişkenlere ln ve karekök dönüşümleri yapıldıktan sonra da normallik testi yapılmış fakat normal dağılmadıkları görülmüştür. Değişkenlerin dönüşümsüz halleriyle devam edilecek ve analizin devamlılığı için değişkenlerin çoklu normal dağıldığı varsayılacaktır.

İkinci varsayım olan varyansların homojenliği için,

H0: Gruplar arası varyans kovaryans matrisleri eşittir

HA: Gruplar arası varyans kovaryans matrisleri eşit değildir.

Box's Test of Equality of Covariance Matrices^a

Box's M	30,065
F	,895
df1	30
df2	2071,424
Sig.	,630

Box’s M testinin sig. değerinden ($0,630 > 0,05$) H0 reddedilemez. Gruplar arası varyanslar homojendir ve analizin bu varsayımı sağlanmıştır.

Between-Subjects Factors

		Value Label	N
slope	0	yukarı eğimli	21
	1	eğimsiz-düz	140
	2	aşağı eğimli	142
kagt	0	tipik angina	143
	1	atipik angina	50
	2	non anginal	87
	3	asemptomatik angina	23

Kalp ağrısı türleri için de gruplar arası gözlem sayılarının oldukça dengesiz olduğu görülmüştür.

Descriptive Statistics

	slope	kagt	Mean	Std. Deviation	N
yas	yukarı eğimli	tipik angina	58,09	6,715	11
		atipik angina	52,00	5,657	2
		non anginal	46,40	7,232	5
		asemptomatik angina	62,67	3,512	3
		Total	55,38	8,316	21
	eğimsiz-düz	tipik angina	56,32	7,931	84
		atipik angina	53,33	7,632	12
		non anginal	56,30	9,761	33
		asemptomatik angina	56,82	9,390	11
		Total	56,10	8,437	140
	aşağı eğimli	tipik angina	54,04	9,126	48
		atipik angina	50,67	10,276	36
		non anginal	52,37	8,897	49
		asemptomatik angina	52,44	11,479	9
		Total	52,51	9,490	142
	Total	tipik angina	55,69	8,313	143
		atipik angina	51,36	9,512	50
		non anginal	53,52	9,420	87
		asemptomatik angina	55,87	10,033	23
		Total	54,37	9,082	303
drgn_kbsnc	yukarı eğimli	tipik angina	143,55	25,224	11
		atipik angina	115,00	7,071	2
		non anginal	129,40	8,173	5
		asemptomatik angina	157,67	17,786	3
		Total	139,48	22,576	21
	eğimsiz-düz	tipik angina	132,56	17,579	84
		atipik angina	130,92	12,079	12
		non anginal	129,88	16,933	33
		asemptomatik angina	136,55	23,101	11
		Total	132,10	17,420	140
	aşağı eğimli	tipik angina	128,44	16,032	48
		atipik angina	128,31	17,071	36
		non anginal	130,82	17,000	49
		asemptomatik angina	140,56	13,258	9
		Total	129,99	16,591	142
	Total	tipik angina	132,02	18,036	143
		atipik angina	128,40	15,837	50
		non anginal	130,38	16,473	87
		asemptomatik angina	140,87	19,573	23
		Total	131,62	17,538	303

Tanımlayıcı istatistikler tablosuna bakıldığında;

En yüksek yaş ortalaması yukarı eğimli ST segmenti ve asemptomatik kalp ağrısına sahip hastalarda görülmüştür.

En düşük yaş ortalaması yukarı eğimli ST segmenti ve anginal olmayan ağrıya sahip hastalarda görülmüştür.

En yüksek durgun kan basıncı ortalaması da yukarı eğimli ST segmenti ve asemptomatik ağrıya sahip hastalarda görülmüştür.

En düşük durgun kan basıncı ortalaması ise yukarı eğimli ST segmenti ve atipik anginaya sahip hastalarda görülmüştür.

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,962	3683,218 ^b	2,000	290,000	,000	,962
	Wilks' Lambda	,038	3683,218 ^b	2,000	290,000	,000	,962
	Hotelling's Trace	25,402	3683,218 ^b	2,000	290,000	,000	,962
	Roy's Largest Root	25,402	3683,218 ^b	2,000	290,000	,000	,962
slope	Pillai's Trace	,022	1,642	4,000	582,000	,162	,011
	Wilks' Lambda	,978	1,642 ^b	4,000	580,000	,162	,011
	Hotelling's Trace	,023	1,642	4,000	578,000	,162	,011
	Roy's Largest Root	,020	2,938 ^c	2,000	291,000	,055	,020
kagt	Pillai's Trace	,061	3,036	6,000	582,000	,006	,030
	Wilks' Lambda	,940	3,051 ^b	6,000	580,000	,006	,031
	Hotelling's Trace	,064	3,067	6,000	578,000	,006	,031
	Roy's Largest Root	,055	5,375 ^c	3,000	291,000	,001	,053
slope * kagt	Pillai's Trace	,048	1,197	12,000	582,000	,281	,024
	Wilks' Lambda	,952	1,194 ^b	12,000	580,000	,283	,024
	Hotelling's Trace	,049	1,191	12,000	578,000	,286	,024
	Roy's Largest Root	,032	1,531 ^c	6,000	291,000	,168	,031

a. Design: Intercept + slope + kagt + slope * kagt

b. Exact statistic

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

H0: Gruplar arasında bağımlı değişkenler açısından anlamlı bir fark yoktur.

HA: Gruplar arasında en az biri açısından anlamlı bir fark vardır.

Çok değişkenli testlerde, tek yönlüde olduğu gibi gruplar arası gözlem sayılarının farklılıkları sebebiyle Pillai's Trace testine bakılır.

Slope için sig. değeri alfadan büyük olduğu için H0 reddedilemez. Bağımlı değişkenler açısından Slope grupları arasında anlamlı bir fark yoktur.

Kagt için sig. değeri alfadan küçük olduğu için H0 reddedilir. Bağımlı değişkenler açısından Kagt grupları arasında anlamlı bir farklılık vardır.

Slope ve Kagt birlikte incelendiğinde ise yine sig. değeri alfadan büyük olduğu için H0 reddedilemez. Bağımlı değişkenler açısından gruplar arası anlamlı fark yoktur.

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
yas	1,321	11	291	,212
drgn_kbsnc	1,302	11	291	,223

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

a. Design: Intercept + slope + kagt + slope * kagt

H0: Gruplar arası varyanslar eşittir.

HA: Gruplar arası varyanslar eşit değildir.

İki değişken için de (sig. değerleri $> \alpha$ oldukları için) H0 reddedilemez. Gruplar arası varyanslar eşittir.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	yas	1938,440 ^a	11	176,222	2,232	,013	,078
	drgn_kbsnc	6255,633 ^b	11	568,694	1,910	,038	,067
Intercept	yas	279054,093	1	279054,093	3534,960	,000	,924
	drgn_kbsnc	1690940,424	1	1690940,424	5679,702	,000	,951
slope	yas	453,780	2	226,890	2,874	,058	,019
	drgn_kbsnc	235,782	2	117,891	,396	,673	,003
kagt	yas	700,370	3	233,457	2,957	,033	,030
	drgn_kbsnc	3877,825	3	1292,608	4,342	,005	,043
slope * kagt	yas	579,408	6	96,568	1,223	,294	,025
	drgn_kbsnc	2400,520	6	400,087	1,344	,238	,027
Error	yas	22971,897	291	78,941			
	drgn_kbsnc	86635,476	291	297,716			
Total	yas	920487,000	303				
	drgn_kbsnc	5342310,000	303				
Corrected Total	yas	24910,337	302				
	drgn_kbsnc	92891,109	302				

a. R Squared = ,078 (Adjusted R Squared = ,043)

b. R Squared = ,067 (Adjusted R Squared = ,032)

H0: Yaş ve durgun kan basıncı açısından slope kategorileri arasında anlamlı bir fark yoktur.

HA: Yaş ve durgun kan basıncı açısından slope kategorileri arasında anlamlı bir fark vardır.

2 değişken için de slope kategorileri arasındaki testlere bakıldığında sig değerlerinden (0,058 ve 0,673) H0 reddedilemez. Yaş ve durgun kan basıncı, ST segmenti eğimini etkilemiyor denebilir.

H0: Yaş ve durgun kan basıncı açısından kagıt kategorileri arasında anlamlı bir fark yoktur.

HA: Yaş ve durgun kan basıncı açısından kagıt kategorileri arasında anlamlı bir fark vardır.

2 değişken için de kagıt kategorileri arasındaki testlere bakıldığında sig değerlerinden (0,033 ve 0,005) H0 reddedilir. Yaş ve kan basıncı, kalp ağrısı türünü etkiliyor denebilir.

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) kagıt	(J) kagıt	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
yas	tipik angina	atipik angina	4,33*	1,460	,017	,56	8,10
		non anginal	2,18	1,208	,275	-,95	5,30
		asemptomatik angina	-,18	1,996	1,000	-5,33	4,98
	atipik angina	tipik angina	-4,33*	1,460	,017	-8,10	-,56
		non anginal	-2,16	1,577	,520	-6,23	1,92
		asemptomatik angina	-4,51	2,239	,185	-10,29	1,27
	non anginal	tipik angina	-2,18	1,208	,275	-5,30	,95
		atipik angina	2,16	1,577	,520	-1,92	6,23
		asemptomatik angina	-2,35	2,083	,672	-7,74	3,03
	asemptomatik angina	tipik angina	,18	1,996	1,000	-4,98	5,33
		atipik angina	4,51	2,239	,185	-1,27	10,29
		non anginal	2,35	2,083	,672	-3,03	7,74
drgn_kbsnc	tipik angina	atipik angina	3,62	2,835	,578	-3,70	10,95
		non anginal	1,64	2,346	,897	-4,42	7,70
		asemptomatik angina	-8,85	3,876	,104	-18,86	1,17
	atipik angina	tipik angina	-3,62	2,835	,578	-10,95	3,70
		non anginal	-1,98	3,062	,917	-9,89	5,93
		asemptomatik angina	-12,47*	4,347	,023	-23,70	-1,24
	non anginal	tipik angina	-1,64	2,346	,897	-7,70	4,42
		atipik angina	1,98	3,062	,917	-5,93	9,89
		asemptomatik angina	-10,49*	4,046	,049	-20,94	-,04
	asemptomatik angina	tipik angina	8,85	3,876	,104	-1,17	18,86
		atipik angina	12,47*	4,347	,023	1,24	23,70
		non anginal	10,49*	4,046	,049	,04	20,94

Based on observed means.

The error term is Mean Square(Error) = 297,716.

*. The mean difference is significant at the ,05 level.

Tek yönlü analizde Slope değişkeni için çoklu karşılaştırma yapıldığı için burada Kagıt için çoklu karşılaştırma yapmak yeterli olacaktır.

Levene testinin sonucunda varyanslar eşit çıktığı için Tukey testine bakmak doğru olacaktır.

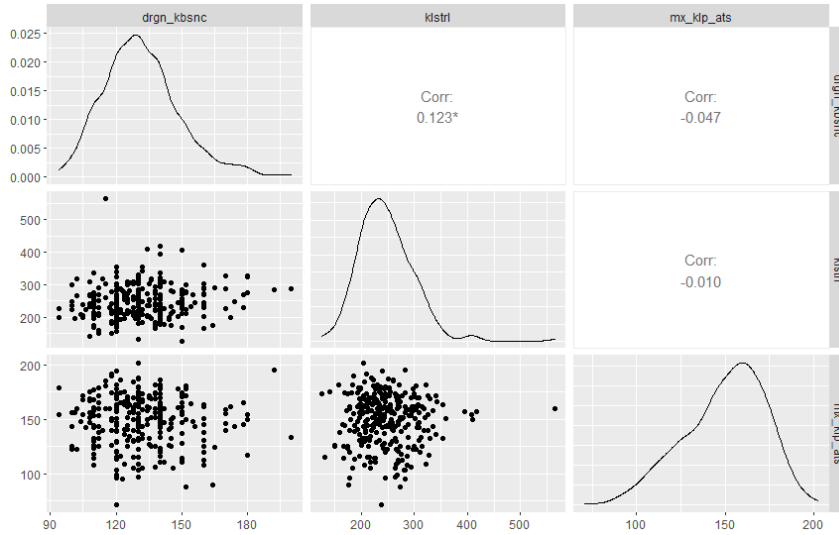
Yaşa bakıldığında, tipik angina ve atipik anginaya sahip hastalar arasında 4,33 birim anlamlı bir fark olduğu tespit edilmiştir. Hem sig. değeri (0,017) < α ; hem de güven aralığının sıfır değerini içermesiyle bu sonuca varılabilmektedir.

Durgun kan basıncına bakıldığında; asemptomatik anginaya sahip hastaların, hem atipik anginal ağrıya sahip hastalarla hem de anginal olmayan ağrıya sahip hastalarla arasında anlamlı bir fark olduğu tespit edilmiştir. Atipik anginaya sahip hastalarla ortalama 12,47 birim, anginal olmayan ağrıya sahip hastalarla ise ortalama 10,49 birim fark tespit edilmiştir.

Yine sig. değerinin alfadan küçük olması ve güven aralıklarının sıfır değerini içermesiyle bu sonuçlara varılabilir.

TEMEL BİLEŞENLER ANALİZİ

Temel bileşenler analizi yapmak için seçtiğim değişkenler drgn_kbsnc (hasta dinlenirken kan basıncı), klstrl (kolesterol) ve mx_klp_ats (maksimum kalp atış hızı).



Korelasyon saçılım matrisine bakıldığında değişkenlerin birbirleriyle ilişkili olmadığı görülür. Değişkenler arası en az ilişki kolesterol ve maksimum kalp atışında, en fazla ilişki ise kolesterol ve durgun kan basıncı arasında çıkmıştır.

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.50348328		
drgn_kbsnc	klstrl	mx_klp_ats
0.50232333	0.50264453	0.51839086

H0: Örneklem, Temel Bileşenler Analizi yapmak için yeterli değildir.

HA: Örneklem, Temel Bileşen Analizi yapmak için yeterlidir.

H0 reddedilir. KMO örneklem yeterlilik ölçütlerine bakıldığında değişkenlerin 0.5 ile sınırda fakat yeterli olduğu görülüyor.

H0: Korelasyon matrisi birim matristir, değişkenler arasında anlamlı bir ilişki yoktur.

HA: Korelasyon matrisi birim matris değildir, değişkenler arasında anlamlı bir ilişki vardır.

```
$chisq
[1] 5.249634

$p.value
[1] 0.1544042

$df
[1] 3
```

Bartlett Test sonucu gösterir ki H_0 reddedilemez. P (sig.) değeri $0.154 > \alpha$ (0.05) olduğundan değişkenler arasında istatistiksel olarak anlamlı bir ilişki yoktur.

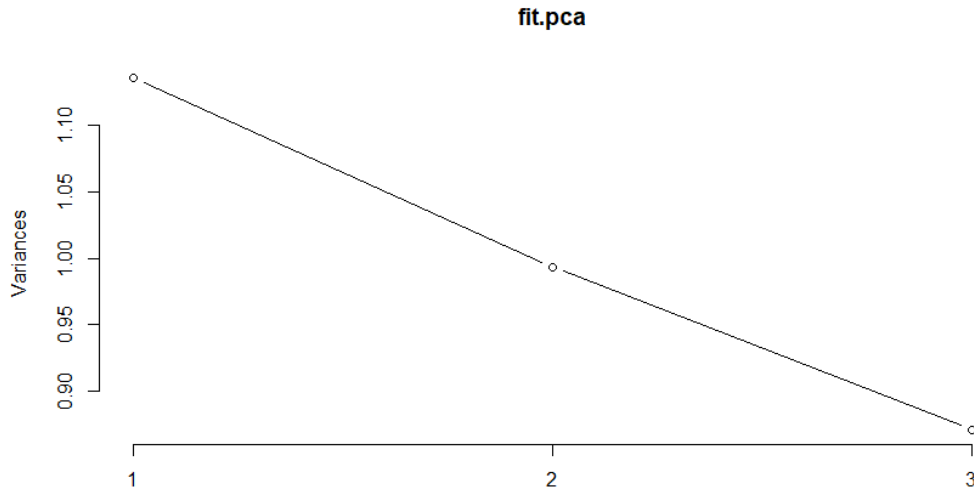
```
> summary(fit.pca)
Importance of components:
               PC1      PC2      PC3
Standard deviation 1.0655 0.9967 0.9334
Proportion of Variance 0.3784 0.3311 0.2904
Cumulative Proportion 0.3784 0.7096 1.0000
> (fit.pca$sdev)^2
[1] 1.1352654 0.9934314 0.8713032
> |
```

Başta değişken sayısı kadar bileşen oluşturuldu. Bileşenlerin özdeğerlerine bakıldığında 1'den büyük olan bileşenlerle devam edilebilirdi fakat burada görülüyor ki 2. Bileşen hem 1. Bileşene'e çok yakın (0.9967) hem de 1. Bileşen toplam varyansın yalnızca %38'ini açıklayabiliyorken 2. Bileşenle birlikte açıklanan toplam varyans %70'tir. Bu yüzden 1. ve 2. Bileşenler ile devam edilecektir.

Factor Pattern			
		Factor1	Factor2
drgn_kbsnc	drgn_kbsnc	0.74285	0.05650
klstrl	klstrl	0.69906	0.35013
mx_klp_ats	mx_klp_ats	-0.30782	0.93148

Variance Explained by Each Factor	
Factor1	Factor2
1.1352654	0.9934314

Değişkenlerin bileşenler tarafından açıklanma oranları verilmiştir. Durgun kan basıncı ve kolesterol 1. Bileşen tarafından açıklanırken 2. Bileşenin maksimum kalp atışını %93 açıkladığı görülmüştür.



Aslında bu grafikte eğimli bir düşüşten sonra grafikte bir kırılım olması beklenir. Bu kırılım noktası bize kaç bileşenle devam etmemiz gerektiğini de gösterirdi fakat bizim değişkenlerimiz arasında anlamlı bir ilişki olmadığından bu grafikteki kırılım oldukça belirsiz.

Bileşenlerin denklemi şu şekildedir.

1. Bileşen = $0,74\text{drgn_kbsnc} + 0,69\text{klstrl} - 0,30\text{mx_klp_ats}$
2. Bileşen = $0,05\text{drgn_kbsnc} + 0,35\text{klstrl} + 0,93\text{mx_klp_ats}$

FAKTÖR ANALİZİ

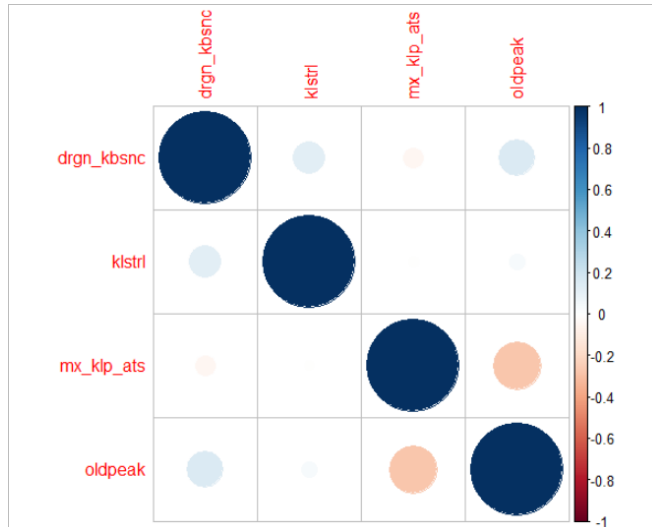
Verideki nicel değişkenlerden yaş hariç tümüyle faktör analizi yapıldı. Yaş değişkeni ise faktör sonuçlarını yorumlarken yardımcı olabileceği düşünülerek analize dahil edilmedi.

Analiz yapılan değişkenler dinlenen hastanın kan basıncı, kolesterol, maksimum kalp atışı ve hasta dinlenirken egzersizin neden olduğu ST depresyonu.

Correlation Matrix^a

		drgn_kbsnc	klstrl	mx_klp_ats	oldpeak
Correlation	drgn_kbsnc	1,000	,123	-,047	,158
	klstrl	,123	1,000	-,010	,031
	mx_klp_ats	-,047	-,010	1,000	-,271
	oldpeak	,158	,031	-,271	1,000
Sig. (1-tailed)	drgn_kbsnc		,016	,209	,003
	klstrl	,016		,432	,297
	mx_klp_ats	,209	,432		,000
	oldpeak	,003	,297	,000	

a. Determinant = ,889



Korelasyon matrisi bize değişkenler arasındaki ilişkiyi veriyor. Matrisin determinantının 0'a uzak olması değişkenler arasındaki ilişki varlığının az olması anlamına geliyor. Teker teker bakıldığında da değişkenlerin birbirleriyle olan korelasyonlarının oldukça sıfıra yakın olduğu görülüyor. En az ilişki maksimum kalp atışı ve kolesterol arasındayken, en fazla ilişkili olan değişkenler maksimum kalp atışı ve egzersize bağlı ST segmentidir.

Inverse of Correlation Matrix

	drgn_kbsnc	klstrl	mx_klp_ats	oldpeak
drgn_kbsnc	1,041	-,123	,004	-,160
klstrl	-,123	1,016	,001	-,011
mx_klp_ats	,004	,001	1,079	,292
oldpeak	-,160	-,011	,292	1,105

Korelasyon matrisinin tersinden varyans şişme değerlerine bakıldığında,

H0: Çoklu bağlantı problemi vardır / VIF değerleri 5'ten küçüktür.

HA: Çoklu bağlantı problemi yoktur / VIF değerleri 5'ten büyüktür.

Ters korelasyon matrisi köşegen elemanlarında 5'ten büyük bir değer olmadığından H0 reddedilir. Çoklu bağlantı problemi yoktur.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	,519
Bartlett's Test of Sphericity	Approx. Chi-Square
	35,143
	df
	6
	Sig.
	,000

KMO ve Bartlett testlerine bakıldığında, KMO test değerinin 0,52 çıktığı görülüyor. Bu verinin örneklem yeterliliğine uygun olmadığı görülür. Burada analize son verilip gözlem

sayısını arttırmaya gidilmesi gerekirdi fakat projenin devamlılığı açısından verinin örneklem yeterliliğine uygun olduğu varsayılmıştır.

Bartlett testi ile kurulan

H0: Değişkenler arasında istatistiksel anlamda bir ilişki yoktur, korelasyon matrisi birim matristir.

HA: Değişkenler arasında istatistiksel anlamda bir ilişki vardır, korelasyon matrisi birim matris değildir.

hipotezi, testin sig. değerinin 0,05'ten küçük olması gereğince H0 reddedilir. Değişkenler arasında istatistiksel olarak anlamlı bir ilişki vardır.

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.51942146			
drgn_kbsnc	klstrl	mx_klp_ats	oldpeak
0.53666590	0.52812036	0.51455782	0.51469497

Değişkenlerin tek tek KMO-MSA değerlerine bakıldığında 0.5'ten küçük olmadıkları, dolayısıyla örneklem yeterliliğine uydukları söylenebilir.

Final Communality Estimates: Total = 2.422924			
drgn_kbsnc	klstrl	mx_klp_ats	oldpeak
0.53246783	0.63007236	0.62957513	0.63080916

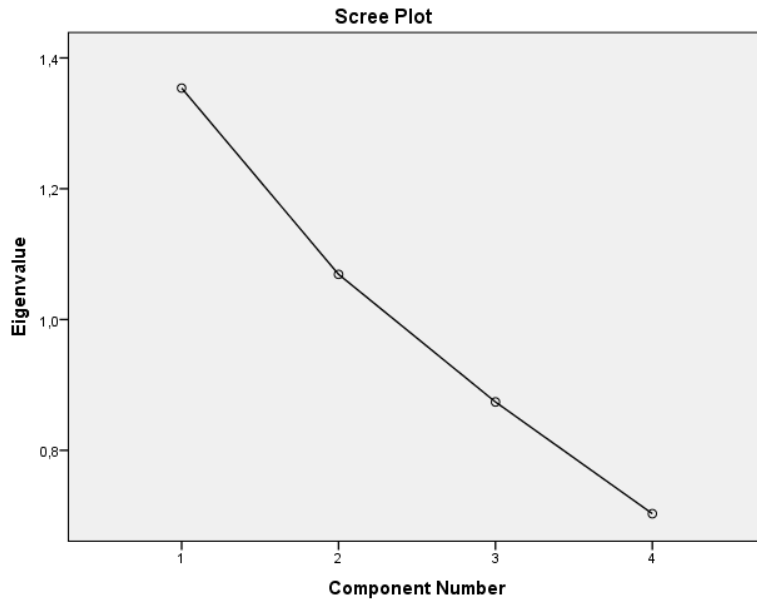
Communalities tablosunda, faktörlerin değişkenleri açıklama oranlarını verilmiştir. Bu oranların 0.5'ten büyük olması değişkenlerin faktör analizinde kalması için gerek ve yeterlidir.

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1,354	33,846	33,846	1,354	33,846	33,846	1,289	32,227	32,227
2	1,069	26,727	60,573	1,069	26,727	60,573	1,134	28,347	60,573
3	,874	21,849	82,422						
4	,703	17,578	100,000						

Extraction Method: Principal Component Analysis.

Varyans açıklama tablosunda faktörlerin toplam varyansı açıklama oranları görülüyor. 1. Faktör %33,84'ünü; 2. Faktör %26,727'sini; 3. Faktör %21,874'ünü ve 4. Faktör %17,578'ini açıklıyor. 1. Ve 2. Faktörler birlikte toplam varyansın %60,573'ünü açıklıyorlar. 4 değişkeni sınıflamak istediğimiz bir faktör analizinde 3 faktöre indirgemek pek de mantıklı olmadığından ve 2 faktör toplam varyansın %60'ını açıkladığından 2 faktöre indirgenerek devam edilecektir.

Ayrıca tablonun sağ tarafında döndürme işleminden sonra 1. Faktörün varyans açıklama oranı %1 puan düşerken 2. Faktörün %2 puan arttığı, kümülatif açıklama oranının değişmediği de görülmüştür.



Faktör numarası ve öz değer grafiğine bakıldığında, grafikteki kırılımın 2. Faktör noktasında olması da 2 faktöre indirgemenin mantıklı olduğunu doğrular.

Component Matrix^a

	Component	
	1	2
drgn_kbsnc	,520	,512
klstrl		,748
mx_klp_ats	-,659	,442
oldpeak	,761	

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Loadings:

	PC1	PC2
drgn_kbsnc	0.520	0.512
mx_klp_ats	-0.659	0.442
oldpeak	0.761	-0.228
klstrl	0.265	0.748

	PC1	PC2
SS loadings	1.354	1.069
Proportion Var	0.338	0.267
Cumulative Var	0.338	0.606

>
> |

Bileşen matris hangi değişkenin hangi faktör tarafından ne kadar açıklandığını gösterir. Bir değişkenin, faktör tarafından %10 açıklanması o faktöre yüklenebilmesi için yeterliken analize başlarken bu oran %30 seçilmiştir. Klstrl ve olpeak değişkenleri sırasıyla 2. ve 1. faktörlere atanabilmiştir fakat drgn_kbsnc değişkeni iki faktör tarafından da neredeyse eşit şekilde açıklanabildiği görülmüştür. Aynı şekilde mx_klp_ats değişkeni de iki faktöre de yüklenmiştir.

Bu fazla yüklenme sorununun çözülmesi için Varimax döndürme yöntemi kullanılmıştır.

Reproduced Correlations

		drgn_kbsnc	klstrl	mx_klp_ats	oldpeak
Reproduced Correlation	drgn_kbsnc	,532 ^a	,521	-,116	,279
	klstrl	,521	,630 ^a	,156	,031
	mx_klp_ats	-,116	,156	,630 ^a	-,602
	oldpeak	,279	,031	-,602	,631 ^a
Residual ^b	drgn_kbsnc		-,398	,070	-,120
	klstrl	-,398		-,166	-,001
	mx_klp_ats	,070	-,166		,331
	oldpeak	-,120	-,001	,331	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 5 (83,0%) nonredundant residuals with absolute values greater than 0.05.

Yeniden üretilmiş korelasyon matrisine bakıldığında değişkenler arasındaki ilişkinin arttığı söylenebilir.

Rotated Component Matrix^a

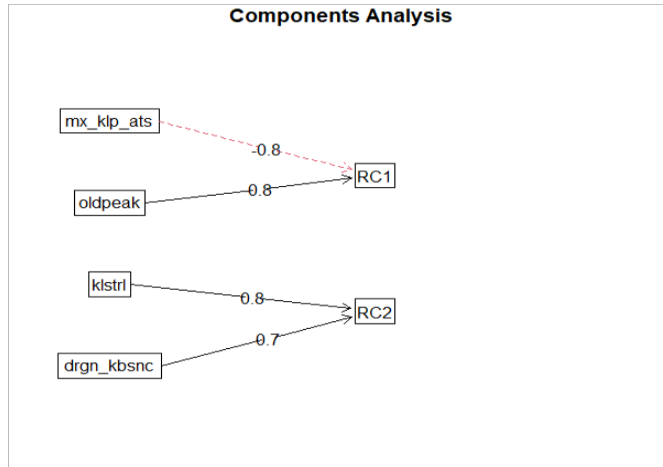
	Component	
	1	2
drgn_kbsnc		,698
klstrl		,784
mx_klp_ats	-,790	
oldpeak	,777	

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

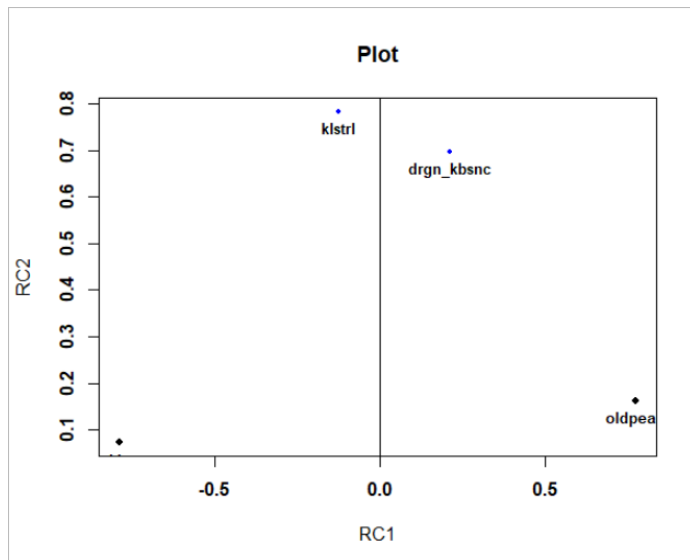
a. Rotation converged in 3 iterations.

Döndürülmüş bileşen matrisine bakıldığında 4 değişkenin de faktörlere fazla yüklenme yapmadan dağılabildiği görülmüştür.



Değişkenler de atandığına göre faktörlerin denklemleri şu şekilde yazılır.

1. faktör: $-0,790mx_klp_ats + 0,777oldpeak$
2. faktör: $0,698drgn_kbsnc + 0,784klstrl$

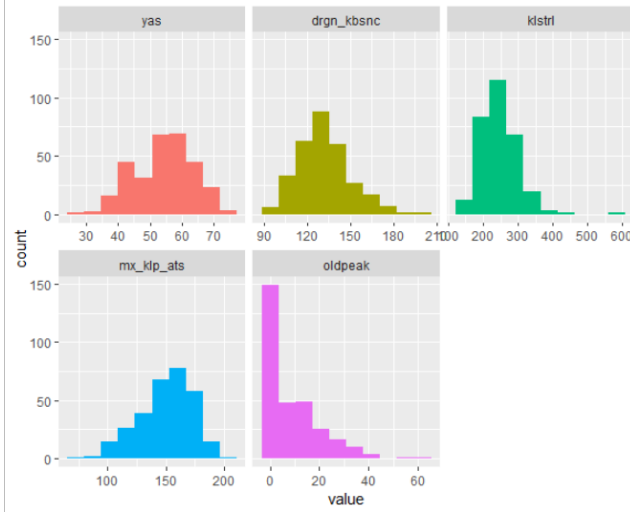


Oldpeak, kan akışı yavaşlayan hastada egzersizin neden olduğu ST depresyonuydu; maksimum kalp atışı, egzersizin değiştirebileceği bir değişken olduğu düşünülerek birinci faktöre “Egzersiz Faktörü” denebilir.

Sağlıklı beslenmeyle kolesterol ve kan basıncı değerleri ortak şekilde artıp azalacağından bu faktöre de “Beslenme Faktörü” denebilir.

DISKRİMİNANT ANALİZİ

Analizde target yani kalp hastası olup olmama durumunu yordamak için bağımlı değişken seçilmiştir. Bağımsız değişken içinse verideki tüm nicel değişkenler alınmıştır. Bunlar yaş, durgun kan basıncı, kolesterol, maksimum kalp atışı ve kan akışı yavaşladığında egzersize bağlı ST depresyonu değişkenleridir.



Değişkenlerin kendi içlerinde dağılımları verilmiştir.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
drgn_kbsnc	,102	303	,000	,966	303	,000
klstl	,055	303	,025	,947	303	,000
yas	,076	303	,000	,986	303	,006
mx_klp_ats	,071	303	,001	,976	303	,000
oldpeak	,216	303	,000	,789	303	,000

a. Lilliefors Significance Correction

Değişkenlere tek tek baktığımızda hiçbirinin normal dağılmadığı görülüyor.

```
+ dplyr::select(!c(target)) %>%
+ mshapiro_test()
# A tibble: 1 x 2
  statistic      p.value
  <dbl>         <dbl>
1    0.943 0.0000000177
> |
```

Teker teker normal dağılmadığı görülen değişkenlerin, çok değişkenli normal dağılıma da uymadıkları R’da yapılan Shapiro-Wilk testiyle tespit edilmiştir. Değişkenlere ln ve karekök dönüşümleri yapılmasına rağmen normallik varsayımı sağlanamamıştır.

Diskriminant Analizi temel varsayımlarından olan çoklu normallik varsayımı sağlanamadığı için bu değişkenler ile analiz yapılamaz fakat projenin devamlılığı için değişkenlerin çoklu normalliğe uyduğu varsayılacaktır.

H0: Bağımlı değişkendeki gruplar arası varyans-kovaryanslar eşittir.

HA: Bağımlı değişkendeki gruplar arası varyans-kovaryanslar eşit değildir.

Test Results

Box's M		91,948
F	Approx.	6,020
	df1	15
	df2	340980,412
	Sig.	,000

Tests null hypothesis of equal population covariance matrices.

Box's M testinin sonucunda sig. değeri $(0,000) < \alpha (0,05)$ olduğundan H0 hipotezi reddedilir. Gruplar arası varyans kovaryanslar eşit olmadığından analize devam edilmemesi gerekir. Projenin devamlılığı için varyans kovaryansların eşit olduğu varsayılacaktır.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,324 ^a	100,0	100,0	,495

a. First 1 canonical discriminant functions were used in the analysis.

Özdeğerler tablosuna gelindiğinde ise kanonik korelasyon değerinin karesinden $(0,245)$ görülmüştür ki bu değişkenlerle yapılan analizin varyans açıklama oranı oldukça düşüktür. Özdeğer değerinin $0,4$ 'ten büyük olması model için gerek ve yeter koşullardan biriydi fakat bu modelde $0,4$ 'ten küçük olduğu görülmüştür. Model, toplam varyansın %25'ini açıklamaktadır.

H0: Diskriminant fonksiyonu önemsizdir.

HA: Diskriminant fonksiyonu önemlidir.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,755	83,831	5	,000

Fonksiyon için Wilks' Lambda testi p değeri (0,000) < α olduğundan H0 reddedilir.

Diskriminant fonksiyonu önemlidir.

Structure Matrix

	Function
	1
mx_klp_ats	-,817
oldpeak	,633
yas	,406
drgn_kbsnc	,257
klstrl	,150

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

Structure matris ise değişkenlerin diskriminant fonksiyonunu en çok maksimum kalp atışı değişkeni açıkladığı görülüyor. Daha sonra oldpeak, yaş, durgun kan basıncı ve kolesterol sırasıyla ilişkilerin kuvveti azalmış.

Canonical Discriminant Function Coefficients

	Function
	1
yas	,001
drgn_kbsnc	,011
klstrl	,003
mx_klp_ats	-,036
oldpeak	,048
(Constant)	2,727

Unstandardized coefficients

Diskriminant fonksiyonu şu şekilde yazılır.

$$Y = 2,727 + 0,001yas + 0,011drgn_kbsnc + 0,003klstrl - 0,036mx_klp_ats + 0,048oldpeak$$

Classification Results^a

			Predicted Group Membership		Total
			sağlıklı	hasta	
Original	Count	target	84	54	138
		hasta	33	132	165
	%	sağlıklı	60,9	39,1	100,0
		hasta	20,0	80,0	100,0

a. 71,3% of original grouped cases correctly classified.

%71,3 doğru atama oranı elde edilmiştir.

Prior Probabilities for Groups

target	Prior	Cases Used in Analysis	
		Unweighted	Weighted
sağlıklı	,455	138	138,000
hasta	,545	165	165,000
Total	1,000	303	303,000

Grupların anakütle içindeki oranlarına bakıldığında ise %45'e %55 oranı görülüyor.

H0: Doğru atama oranı yeterli değildir.

HA: Doğru atama oranı yeterlidir.

Bu hipoteze karar vermek için prior oranlarının kareler toplamına bakılmalıdır.

$$0,455^2 + 0,545^2 = 0,50405$$

Doğru atama oranı, bu kareler toplamından büyük olduğundan H0 reddedilir.

Doğru atama oranı yeterlidir.

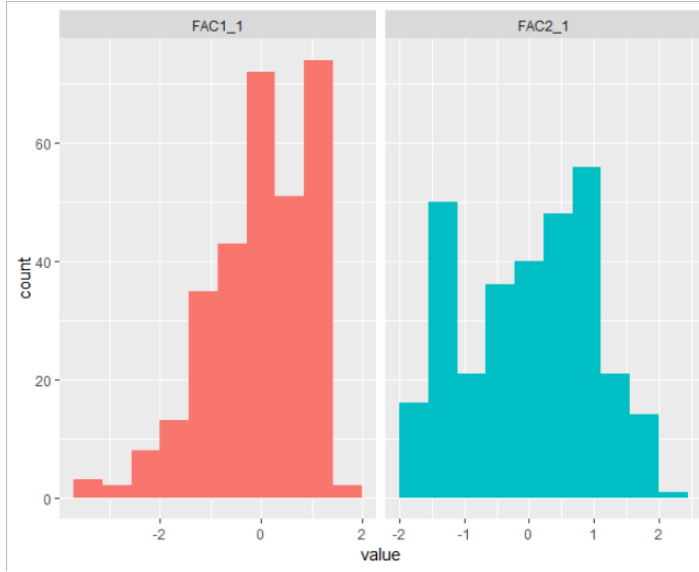
2. DİSKRİMİNANT ANALİZİ 2

Birinci Diskriminant analizinde varyans açıklama oranı çok düşük olduğundan faktör analizinden elde edilen faktör skorlarıyla 2. Diskriminant analizi yapılmaya karar verilmiştir. Bağımlı değişken ise tekrar target (kalp hastası olup olmama durumu) seçilmiştir.

Correlations

			REGR factor score 1 for analysis 1	REGR factor score 2 for analysis 1
Spearman's rho	REGR factor score 1 for analysis 1	Correlation Coefficient	1,000	,002
		Sig. (2-tailed)	.	,975
		N	303	303
	REGR factor score 2 for analysis 1	Correlation Coefficient	,002	1,000
		Sig. (2-tailed)	,975	.
		N	303	303

Faktörler arasındaki korelasyon yok denebilir. Diskriminant analizi yapmaya uygun olduğu görülmüştür.



Faktörlerin kendi içlerinde dağılımları şu şekildedir.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
REGR factor score 1 for analysis 1	,084	303	,000	,942	303	,000
REGR factor score 2 for analysis 1	,092	303	,000	,962	303	,000

a. Lilliefors Significance Correction

Değişkenlerin normalliği test edilmiş fakat hem Kolmogorov-Smirnov hem de Shapiro-Wilk testleri ile değişkenlerin normal dağılmadıkları görülmüştür.

```
+ mshapiro_test()
# A tibble: 1 x 2
  statistic    p.value
  <dbl>      <dbl>
1    0.954 0.000000394
> |
```

Yine değişkenlerin çoklu normal dağılım varsayımına uymadığı görülmüştür. Faktör skorlarına da ln ve karekök dönüşümleri yapılmış fakat normallik varsayımı Shapiro-Wilk testine göre sağlanamamıştır. Analize son verilmesi gerekir fakat projenin devamlılığı için normal dağıldığı varsayılacaktır.

H0: Gruplar arası varyans kovaryanslar eşittir.

HA: Gruplar arası varyans kovaryanslar eşit değildir.

Box's M-test for Homogeneity of Covariance Matrices

data: veri[, 2:3]

Chi-Sq (approx.) = 22.291, df = 3, p-value = 5.674e-05

> |

Box's M testinin sig. değerinden (0,05674) H0 reddedilemez. Gruplar arası varyans kovaryanslar homojendir.

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,584 ^a	100,0	100,0	,607

a. First 1 canonical discriminant functions were used in the analysis.

Özdeğer değeri bu defa 0,4'ten büyük çıkmıştır ki bu modelin varyans açıklama oranının önceki modele göre daha yüksek olduğunu gösterir. Kanonik korelasyon değerinin karesine bakıldığında modelin, toplam varyansın %37'sini açıkladığı görülür.

H0: Diskriminant fonksiyonu önemsizdir.

HA: Diskriminant fonksiyonu önemlidir.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,631	138,021	2	,000

Wilks' Lambda testi sig. değerinden H0 reddedilir. Diskriminant fonksiyonu önemlidir.

Structure Matrix

	Function 1
REGR factor score 2 for analysis 1	,772
REGR factor score 1 for analysis 1	,461

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
Variables ordered by absolute size of correlation within function.

Yapı matrisine bakıldığında Beslenme faktörünün, diskriminant denklemini Egzersiz faktörüne göre daha fazla açıkladığı görülmüştür.

Raw Canonical Coefficients		
Variable	Label	Can1
FAC1_1	REGR factor score 1 for analysis 1	0.688240480
FAC2_1	REGR factor score 2 for analysis 1	1.051314475

Diskriminant fonksiyonu şöyle yazılır.

$$Y = 0,688 \text{egzersiz} + 1,051 \text{beslenme}$$

Number of Observations and Percent Classified into target			
From target	hasta	sağlıklı	Total
hasta	140 84.85	25 15.15	165 100.00
sağlıklı	43 31.16	95 68.84	138 100.00
Total	183 60.40	120 39.60	303 100.00
Priors	0.54455	0.45545	

Bu modelin doğru atama oranı köşegenlerdeki doğru atamaların toplamının (140+95), gözlem sayısına (303) oranından $(140+95) \cdot 100 / 303 = \%77,6$ olarak hesaplanır.

H₀: Doğru atama oranı yetersizdir.

H_A: Doğru atama oranı yeterlidir.

Class Level Information					
target	Variable Name	Frequency	Weight	Proportion	Prior Probability
hasta	hasta	165	165.0000	0.544554	0.544554
sağlıklı	sağlıklı	138	138.0000	0.455446	0.455446

Bu tabloda ise prior probability sütunu, grupların ana kitle içindeki oranlarını vermektedir. Bu oranların kareler toplamının doğru atama oranından küçük olması doğru atama oranının yeterli olduğunu gösterir.

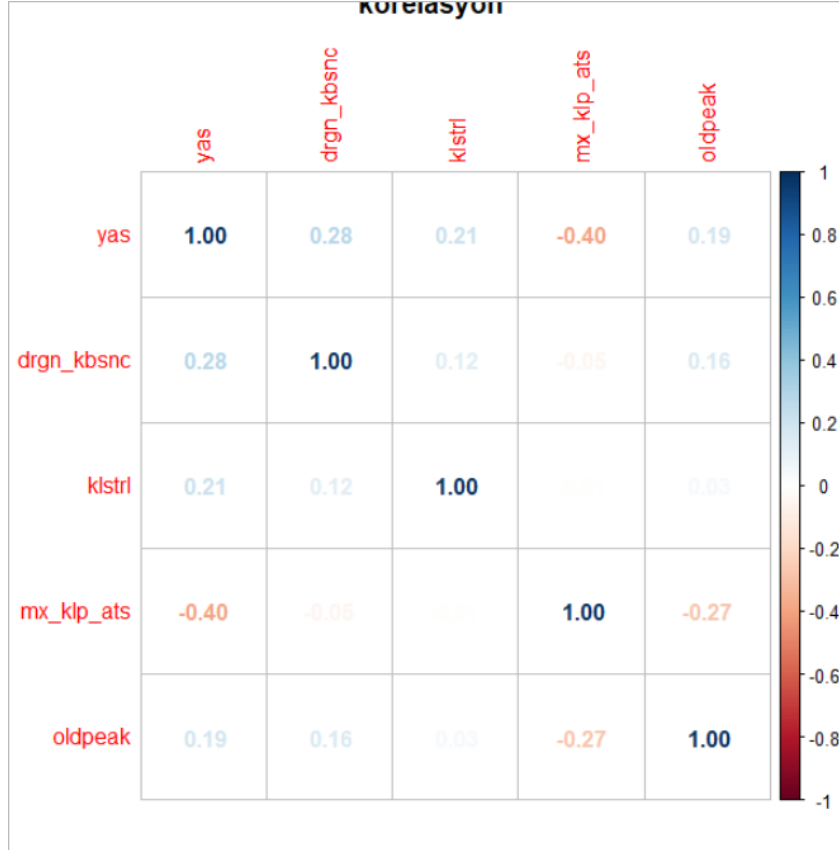
$$0,455^2 + 0,545^2 = 0,50405$$

$0,50 < 0,77$ olduğundan H₀ reddedilir. Doğru atama oranı yeterlidir.

2. Diskriminant analizinin, doğru atama oranlarını kıyaslarsak, 1. Analizden daha başarılı sonuç verdiği görülmüştür.

LOGISTIC REGRESYON ANALİZİ

Bağımlı değişken target (hasta olup olmama durumu) alınmıştır. Bağımsız değişkenler ise kategorik olmayan tüm değişkenlerdir. Kategorik değişkenlerin eklendiği bir model kurulmuş fakat doğru atama oranının değişmediği görülmüştür. Bu yüzden raporda, kategorik değişkenlerin olmadığı modelin sonuçları olacaktır.



Değişkenler arası korelasyona bakıldığında en yüksek korelasyon maksimum kalp atışı ve yaş arasında görülmüştür. Yüksek korelasyon görülmediği için değişken çıkarılmasına gerek yoktur.

> invkor

```
[1,] 1.35264091 -0.31495314 -0.24340442 0.50601283 -0.05870206
[2,] -0.31495314 1.11403750 -0.06656422 -0.11382340 -0.14629756
[3,] -0.24340442 -0.06656422 1.05933929 -0.08977989 -0.00073447
[4,] 0.50601283 -0.11382340 -0.08977989 1.26867285 0.27004416
[5,] -0.05870206 -0.14629756 -0.00073447 0.27004416 1.10740831
```

Yukarıda korelasyon matrisinin tersi verilmiştir.

H0: Çoklu bağlantı problemi yoktur / VIF değerleri 5'ten küçüktür.

HA: Çoklu bağlantı problemi vardır / VIF değerleri 5'ten büyüktür.

Ters korelasyon matrisi köşegen elemanlarında 5'ten büyük bir değer olmadığından H0 reddedilemez. Çoklu bağlantı problemi yoktur.

Dependent Variable Encoding

Original Value	Internal Value
sağlıklı	0
hasta	1

Bağımlı değişkendeki kategorilerde 0 sağlıklı bireyleri, 1 ise hastaları ifade ettiği raporun en başında da belirtilmiştir. Bu analizde de kategoriler belirtildiği gibi alınmıştır. Dolayısıyla analizde hasta olma durumundaki değişim incelenecektir. Analizde Forward stepwise yöntem kullanılacaktır.

Iteration History^{a,b,c,d}

Iteration		-2 Log likelihood	Coefficients		
			Constant	mx_klp_ats	oldpeak
Step 1	1	360,538	-5,320	,037	
	2	359,262	-6,323	,043	
	3	359,257	-6,391	,044	
	4	359,257	-6,391	,044	
Step 2	1	342,185	-4,081	,031	-,044
	2	339,646	-4,957	,038	-,056
	3	339,623	-5,047	,038	-,057
	4	339,623	-5,048	,038	-,057

a. Method: Forward Stepwise (Likelihood Ratio)

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 417,638

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

İterasyon tablosuna baktığımızda, 1. Adımda maksimum kalp atışı değişkeninin modele girdiğini ve modelin açıklanamayan kısmı olan -2 Log likelihood değerini 1,281 birim düşürdüğü görülmüştür. Daha sonra 2. Adımda ise ST depresyonu değerinin (oldpeak) modele dahil olmuş ve -2 Log likelihood değerini toplamda bu iki değişken 20,915 birim düşürmüştür. Diğer değişkenler ise modele anlamlı bir katkısı olmadığı için stepwise yöntem dahilinde modele alınmamıştır.

H0: Modele eklenen bağımsız değişkenler (mx_klp_ats ve oldpeak) modele anlamlı katkı sağlamaz.

HA: Modele eklenen bağımsız değişkenler modele anlamlı katkı sağlar.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	58,381	1	,000
	Block	58,381	1	,000
	Model	58,381	1	,000
Step 2	Step	19,634	1	,000
	Block	78,015	2	,000
	Model	78,015	2	,000

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	mx_klp_ats	1	1	53.8932	<.0001	mx_klp_ats
2	oldpeak	1	2	18.6165	<.0001	oldpeak

İki adımda da modele eklenen bağımsız değişkenlerin, modelin bağımlı değişkeni açıklama oranını arttırdığı Omnibus Testiyle doğrulanmıştır. H₀, testin sig. değerininin 0,05'ten küçük olmasıyla reddedilir.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	359,257 ^a	,175	,234
2	339,623 ^a	,227	,303

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Modelin yapay belirtme katsayılarının oldukça düşük olduğu görülmüştür. Bu yapay belirtme katsayıları başka modellerle kıyaslarırken kullanılır. 2. Adımda bu katsayıların yükselmesi ise yine ikinci eklenen değişkenin modelin açıklamasını yükselttiğinin başka bir göstergesi denebilir.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6,274	8	,617
2	12,029	8	,150

H₀: Gözlenen ve beklenen değerler birbirine yakındır/Model verilere uyumludur.

H_A: Gözlenen ve beklenen değerler birbirine yakın değildir/Model verilere uyumlu değildir.

Hosmer-Lemeshow testinin, 2 adımda da sig. değerinden büyük olması nedeniyle H₀ reddedilemez. Model verilere uyumludur.

The FREQ Procedure				
Frequency Percent Row Pct Col Pct	Table of target by pred			
	target(target)	pred		Total
		0	1	
	0	85 28.05 61.59 74.56	53 17.49 38.41 28.04	138 45.54
	1	29 9.57 17.58 25.44	136 44.88 82.42 71.96	165 54.46
	Total	114 37.62	189 62.38	303 100.00

Sınıflama tablosunda ise doğru atama oranının, 2. Adım sonunda $(\frac{85+136}{303}) = \%72,9$ olduğu görülmüştür.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept	1	-5.0484	1.0145	24.7644	<.0001	0.006
mx_klp_ats	1	0.0382	0.00660	33.5233	<.0001	1.039
oldpeak	1	-0.0574	0.0139	17.1069	<.0001	0.944

Lojistik regresyon modeli denklemi şu şekilde yazılır:

$$\ln\left(\frac{p}{1-p}\right) = -5,048 + 0,038 \text{ mx_klp_ats} - 0,057 \text{ oldpeak}$$

Logistic regresyon analizinde değişkenlerin etkilerini yorumlayabilmek için denklemin her iki tarafına da exponential işlemi yapılır ki bu işlem sonucunda değişkenlerin etkileri tablonun en sağdaki sütununda [Exp(Est)] verilmiştir.

Maksimum kalp atışındaki 1 birimlik artış, kişinin hasta olma olasılığını 1,039 birim arttırır ya da maksimum kalp atışındaki 1 birimlik artış, kişinin hasta olma olasılığını yaklaşık %4 (%039) arttırır denebilir.

ST depresyonu değerindeki 1 birimlik artış, kişinin hasta olma olasılığını 0,944 birim azaltır ya da ST depresyonu değerindeki 1 birimlik artış, kişinin hasta olma olasılığını yaklaşık %6(%056) azaltır denebilir.

KÜMELEME ANALİZİ

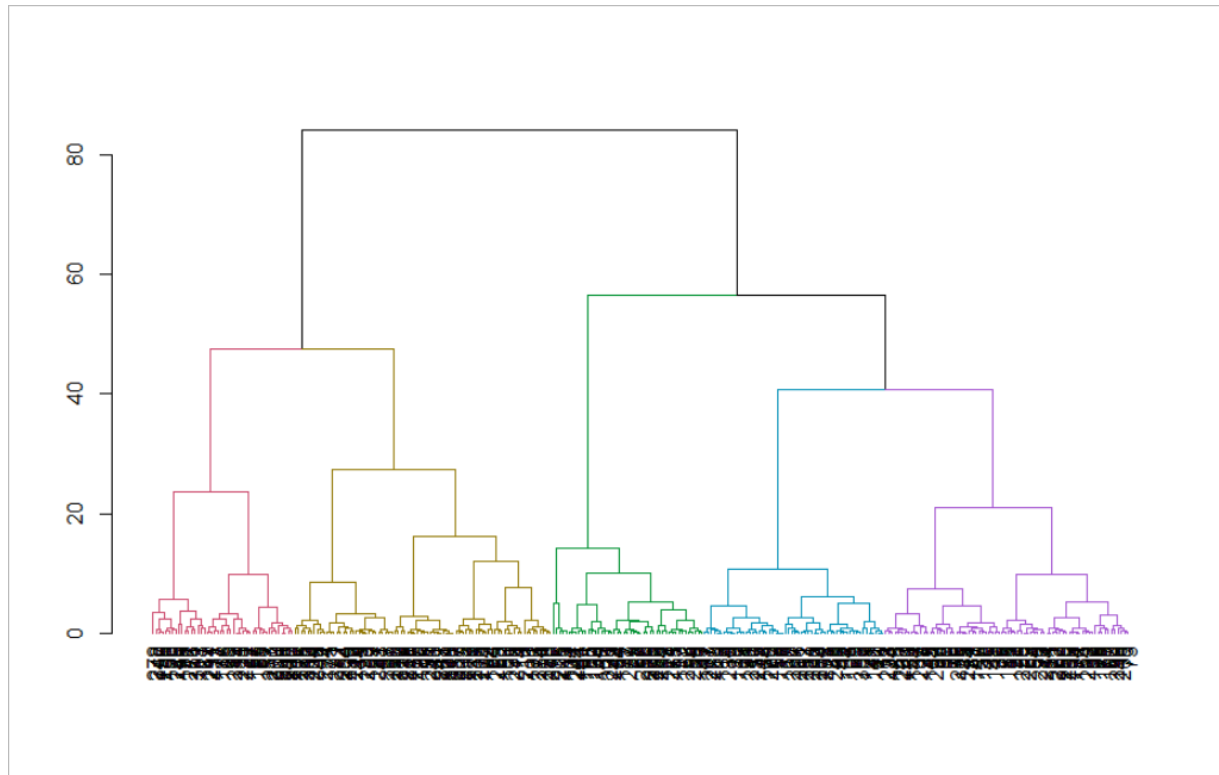
Verideki nicel değişkenlerden oldpeak değişkeninde st depresyonu durumu olmadığında 0 değeri girildiğinden bu değişken kullanılmayacaktır.

```
> rcorr(as.matrix(veri[,-1]),type="pearson")
      yas drgn_kbsnc klstrl mx_klp_ats
yas      1.00      0.28   0.21   -0.40
drgn_kbsnc 0.28      1.00   0.12   -0.05
klstrl     0.21      0.12   1.00   -0.01
mx_klp_ats -0.40     -0.05  -0.01    1.00

n= 303
```

Geriye kalan nicel değişkenlerin korelasyon matrisi verilmiştir. Matriste görüldüğü üzere değişkenler arasında yüksek korelasyon yoktur.

Değişkenlerin birimleri arasındaki farklar göz önünde bulundurularak değişkenlerde standartlaştırma işlemi yapılmış, bu dönüşümün ardından analize başlanmıştır.



Hiyerarşik yöntemlerde Ward metoduyla kümeleme yapıldığında ortaya çıkan dendrogram bu şekildedir. Bu dendrograma bakılarak gözlemlerin 5 kümeye ayrılmasına karar verilmiştir.

K=5 alınarak K-Means yöntemiyle devam edilecektir.

Hiyerarşik yöntem ve Ward metoduyla yapılan ilk kümelemelerde elde edilen küme atamaları ve değişkenlerin ortalamaları, K-Means yönteminin başlangıç küme merkezleri seçilmiştir. Aşağıdaki tablo k-means yönteminin başlangıç küme merkezlerini göstermektedir.

Initial Cluster Centers					
	Cluster				
	1	2	3	4	5
Zscore(yas)	-,26245	,27032	-,13401	-,24403	,20842
Zscore(drgn_kbsnc)	-,16238	,25496	-,24492	,09985	,00773
Zscore(klstrl)	-,12718	,16625	,03724	-,19899	-,02081
Zscore(mx_klp_ats)	,34135	-,83332	,24479	,68339	,11971

Input from FILE Subcommand

Daha sonra analize başlanmış ve 13 iterasyonla kümeler oluşturulmuştur. İterasyon adımları şu şekildedir.

Iteration History ^a					
Iteration	Change in Cluster Centers				
	1	2	3	4	5
1	1,127	,507	,800	,671	,711
2	,215	,323	,348	,116	,190
3	,063	,276	,258	,046	,211
4	,072	,234	,134	,132	,280
5	,063	,196	,094	,105	,203
6	,038	,082	,104	,047	,108
7	,066	,104	,125	,086	,149
8	,086	,072	,107	,043	,101
9	,109	,016	,075	,075	,040
10	,066	,000	,046	,041	,061
11	,029	,050	,039	,038	,000
12	,000	,000	,000	,021	,028
13	,000	,000	,000	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 13. The minimum distance between initial centers is ,244.

Final Cluster Centers					
	Cluster				
	1	2	3	4	5
Zscore(yas)	-,77683	,70024	,31157	-,95225	,75400
Zscore(drgn_kbsnc)	-1,12145	-,21912	-,06799	-,10428	1,39397
Zscore(klstrl)	-,61134	-,28604	1,40788	-,25171	-,06745
Zscore(mx_klp_ats)	-,00787	-1,14601	,17892	,91377	,06453

Final küme merkezleri tablosu verilmiştir. Buna göre 4. Kümenin en genç küme olduğu söylenebilir. Durgun kan basıncının en düşük olduğu küme 1. kümeyken, 5. Küme ise en yüksek kan basıncına sahip küme olmuştur. 3. Küme ise kolesterol değeri en yüksek

kümedir. Kolesterolü en düşük küme ise 1. küme olmuştur. 4. Küme kalp atış hızı en yüksek olan kümeyken, kalp atışı en düşük olan küme ise 2. kümedir.

Final küme merkezlerine bakıldığında

H0: Değişkenler, kümeleme analizi için anlamlı farklılık sağlamaz.

HA: Değişkenler, kümeleme analizi için anlamlı farklılık sağlar.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore(yas)	42,617	4	,441	298	96,552	,000
Zscore(drgn_kbsnc)	42,500	4	,443	298	95,947	,000
Zscore(klstrl)	32,269	4	,580	298	55,609	,000
Zscore(mx_klp_ats)	40,718	4	,467	298	87,213	,000

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

Anova tablosuna bakıldığında tüm değişkenlerin sig değerlerinin analizin hata payından düşük olduğunu görülmüştür. H0 reddedilir. Değişkenler, kümeleme analizi için anlamlı farklılık sağlar ve analizde etkilidir.

KAYNAKÇA

<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

<https://www.kaggle.com/ronitf/heart-disease-uci>