



FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

UYKU RAHATSIZLIĞINA İLİŞKİN VERİ ANALİZİ

ÇOK DEĞİŞKENLİ İSTATİSTİKSEL
YÖNTEMLER
DÖNEM SONU PROJESİ

H.Simay ÖZGÜL – 20201101061
Ataman Önoğlu – 20201101058

İÇİNDEKİLER

VERİ SETİNİN TANIMI	2
MANOVA	3
FAKTÖR ANALİZİ	9
DISKRİMİNANT ANALİZİ	13
LOJİSTİK REGRESYON	16
KÜMELEME	19

VERİ SETİNİN TANIMI

Toplamda 374 gözlem ve 13 değişkenden oluşan veri setinde kayıp gözlem bulunmamaktadır. Aşağıda veri setinin linki verilmiş ve değişkenler açıklanmıştır. Veri setinden hareketle; kişilerin sahip olduğu stres seviyesi, uyku kalitesi ve uyku sürelerinin uyku bozukluğunda ne kadar etkili olduğu ele alınacaktır.

- <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>

PERSON ID: Gözlem Numaraları

GENDER: Cinsiyet

AGE: Yaş

OCCUPATION: Meslek

SLEEP DURATION: Saat cinsinden uyku süresi

QUALITY OF SLEEP: 1 ile 10 arasında ölçülmüş uyku kalitesi

PHYSICAL ACTIVITY LEVEL : Fiziksel aktivite seviyesi (dakika/gün)

STRESS LEVEL: 1 ile 10 arasında ölçülmüş stres seviyesi

BMI CATEGORY: Kişiye ait vücut kitle indeksi

BLOOD PRESSURE: Kişinin kan basıncı ölçümü (sistolik/diastolik)

HEART RATE: Kişinin dinlenme halindeki dakika başına kalp atış hızı

DAILY STEPS: Günlük atılan adım sayısı

SLEEP DISORDER: Uyku bozukluğu durumu

Descriptive Statistics									
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Age	374	27	59	42,18	8,673	,257	,126	-,910	,252
SleepDuration	374	5,8	8,5	7,132	,7957	,038	,126	-1,287	,252
QualityofSleep	374	4	9	7,31	1,197	-,207	,126	-,748	,252
PhysicalActivityLevel	374	30	90	59,17	20,831	,074	,126	-1,266	,252
StressLevel	374	3	8	5,39	1,775	,154	,126	-1,327	,252
HeartRate	374	65	86	70,17	4,136	1,225	,126	2,286	,252
DailySteps	374	3000	10000	6816,84	1617,916	,178	,126	-,394	,252
Valid N (listwise)	374								

H0: Veri seti normal dağılmaktadır.

H1: Veri seti normal dağılmamaktadır.

Yukarıdaki tabloya bakıldığında veri setinde kayıp gözlem olmadığı görülmektedir. Kısaca yorumlamak istersek; veri setindeki gözlemlere ait bilgiler için ortalama yaşın yaklaşık 42, günlük uyku saatinin ortalama 7 saat, uyku kalitesi 7 bandında seyderken stres seviyesinin ortalama 5 seviyelerinde olduğu söylenebilir.

Yine aynı tabloda “Skewness” ve “Kurtosis” değerleri göze çarpmaktadır. Kısaca “Skewness” dağılımın simetrisi, “Kurtosis” ise dağılımın basıklığı hakkında bilgi verir. “Skewness” ile “Kurtosis” değerleri, hata değerleriyle oranlanır ve “+2 ile -2” aralığında çıkan oranların normalliği bozmadığı söylenebilir.

Bahsedilen işlem yapıldığı takdirde oranlar genel olarak bahsedilen aralıktan büyük çıktığı için veri setinin normal dağılmadığı söylenebilir. Dolayısıyla H_0 reddedilir.

MANOVA

Veri seti için denklem modeli “**SleepDuration+ QualityofSleep+ StressLevel = SleepDisorder+ Σ** ” şeklinde önümüze gelmektedir.

Bu teste başlamadan önce bakmamız gereken birkaç varsayımı ele almak gerekmektedir. Aşağıda kısaca bu varsayımlar özetlenmiştir. Analize de varsayım testleri neticesinde devam edilmiştir.

- Grup gözlem sayısı, değişken sayısından büyük olmalı veya sayı minimum 20 olmalıdır.
- Çoklu bağlantı olmamalıdır.
- En büyük gözlem sayısının en küçük gözlem sayısına bölümü “1.5” değerinden küçük olmalıdır. Bu varsayım ortak (homojen) varyans-kovaryans matrisinden bahseder.

Between-Subjects Factors

		N
SleepDisorder	Insomnia	77
	None	219
	Sleep Apnea	78

Descriptive Statistics

	SleepDisorder	Mean	Std. Deviation	N
SleepDuration	Insomnia	6,590	,3872	77
	None	7,358	,7323	219
	Sleep Apnea	7,032	,9748	78
	Total	7,132	,7957	374
QualityofSleep	Insomnia	6,53	,804	77
	None	7,63	,975	219
	Sleep Apnea	7,21	1,646	78
	Total	7,31	1,197	374
StressLevel	Insomnia	5,87	1,463	77
	None	5,11	1,591	219
	Sleep Apnea	5,67	2,334	78
	Total	5,39	1,775	374

MANOVA için seçilmiş olan değişkenler yukarıdaki tabloda detaylı olarak incelenmiştir.

Box's Test of Equality of Covariance Matrices^a

Box's M	246,166
F	20,201
df1	12
df2	213162,582
Sig.	,000

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

a. Design:
Intercept +
SleepDisorder

“Box’s Test of Equality of Covariance Matrices” tablosu bize veri setinin varyans-kovaryans matrisini vermekte olup, yorumlanmasına olanak sağlamaktadır. Bunun için hipotez testi kuralım;

H₀: Gruplara göre varyans-kovaryans matrisi homojendir.

H₁: Gruplara göre varyans-kovaryans matrisi homojen değildir.

Teste ilişkin hipotez testini yorumlayabilmek için tablonun aşağısında bulunan “Significant”, bir diğer adıyla p-value değerine bakmak gerekmektedir.

Tablo değeri 0.05’ten küçük olduğu için H₀ reddedilir. %95 güven düzeyinde gruplara göre varyans-kovaryans matrisi homojen değil, heterojendir diyebiliriz.

Fakat bu durum minimum etkiye sahip olduğundan varsayım sağlandığı düşünülerek analize devam edilmiştir.

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,998	63060,507 ^b	3,000	369,000	,000	,998
	Wilks' Lambda	,002	63060,507 ^b	3,000	369,000	,000	,998
	Hotelling's Trace	512,687	63060,507 ^b	3,000	369,000	,000	,998
	Roy's Largest Root	512,687	63060,507 ^b	3,000	369,000	,000	,998
SleepDisorder	Pillai's Trace	,290	20,928	6,000	740,000	,000	,145
	Wilks' Lambda	,713	22,645 ^b	6,000	738,000	,000	,155
	Hotelling's Trace	,397	24,374	6,000	736,000	,000	,166
	Roy's Largest Root	,385	47,508 ^c	3,000	370,000	,000	,278

a. Design: Intercept + SleepDisorder

b. Exact statistic

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

Yukarıdaki tablo ise bize MANOVA testinin sonuçlarını vermektedir. Bu tablo için ele alınan konu adına farklı bir hipotez testi aşağıdaki şekilde kurulabilir.

H₀: “SleepDisorder” gruplarına göre bağımlı değişken uygulamaları eşittir.

H₁: “SleepDisorder” gruplarına göre bağımlı değişkenlerden en az biri farklılık göstermektedir.

Tüm varsayımlar sağlanıyor olsaydı “Roy’s Largest Root” test değerlerine bakılması gerekirdi ancak veri setinde temel varsayımlar sağlandığından “Multivariate Tests” tablosunda “Wilk’s Lambda” değerlerine bakılması uygun olacaktır.

Bu test için p-value değerinin 0.05’ten küçük olması nedeniyle H_0 hipotezi reddedilir. %95 güven düzeyinde bağımsız değişkendeki gruplara göre bağımlı değişken uygulamalarından en az biri farklılık göstermektedir.

Aynı zamanda bağımsız değişken, bağımlı değişkenlerin etki büyüklüğünün %15.5’ini açıklamaktadır.

Levene's Test of Equality of Error Variances ^a					
		Levene Statistic	df1	df2	Sig.
SleepDuration	Based on Mean	65,501	2	371	,000
	Based on Median	60,250	2	371	,000
	Based on Median and with adjusted df	60,250	2	352,909	,000
	Based on trimmed mean	66,488	2	371	,000
QualityofSleep	Based on Mean	67,753	2	371	,000
	Based on Median	23,132	2	371	,000
	Based on Median and with adjusted df	23,132	2	251,009	,000
	Based on trimmed mean	69,560	2	371	,000
StressLevel	Based on Mean	43,375	2	371	,000
	Based on Median	19,933	2	371	,000
	Based on Median and with adjusted df	19,933	2	322,836	,000
	Based on trimmed mean	42,840	2	371	,000

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

a. Design: Intercept + SleepDisorder

Yukarıdaki tabloda anlatılmak istenen, varyansın homojen olup olmadığını göstermektir. Varyansın homojen olmadığı testler için “Tamhane”, homojen olduğu testler için “Tukey” testine bakılması gerekir.

Veri setinde incelenen bütün gruplar için p-value değeri 0.05’ten küçük olduğu için varyansların homojen olmadığı söylenebilir. Dolayısıyla “Multiple Comparisons” tablosunda “Tamhane” testine bakılacaktır.

Tests of Between-Subjects Effects							
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	SleepDuration	34,662 ^a	2	17,331	31,913	,000	,147
	QualityofSleep	69,215 ^b	2	34,607	27,601	,000	,130
	StressLevel	40,375 ^c	2	20,188	6,604	,002	,034
Intercept	SleepDuration	14491,631	1	14491,631	26685,381	,000	,986
	QualityofSleep	15025,642	1	15025,642	11983,469	,000	,970
	StressLevel	9128,090	1	9128,090	2985,875	,000	,889
SleepDisorder	SleepDuration	34,662	2	17,331	31,913	,000	,147
	QualityofSleep	69,215	2	34,607	27,601	,000	,130
	StressLevel	40,375	2	20,188	6,604	,002	,034
Error	SleepDuration	201,473	371	,543			
	QualityofSleep	465,184	371	1,254			
	StressLevel	1134,181	371	3,057			
Total	SleepDuration	19260,260	374				
	QualityofSleep	20535,000	374				
	StressLevel	12020,000	374				
Corrected Total	SleepDuration	236,135	373				
	QualityofSleep	534,398	373				
	StressLevel	1174,556	373				

a. R Squared = ,147 (Adjusted R Squared = ,142)

b. R Squared = ,130 (Adjusted R Squared = ,125)

c. R Squared = ,034 (Adjusted R Squared = ,029)

“Test of Between-Subject Effects” tablosu “SleepDisorder” değişkeninin, diğer değişkenler üzerinde anlamlı olup olmadığını vermektedir. Bunun için benzer şekilde p-value değerlerini incelenmesi gerekmektedir.

Bu değerlerin hepsi 0.05ten küçük olduğu için tüm bağımlı değişkenlerde grup ortalamaları anlamlı farklılık göstermektedir.

Multiple Comparisons								
Dependent Variable		(I) SleepDisorder	(J) SleepDisorder	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
SleepDuration	Tukey HSD	Insomnia	None	-.769*	.0976	.000	-.999	-.539
			Sleep Apnea	-.442*	.1184	.001	-.721	-.164
		None	Insomnia	.769*	.0976	.000	.539	.999
			Sleep Apnea	.326*	.0972	.002	.098	.555
		Sleep Apnea	Insomnia	.442*	.1184	.001	.164	.721
			None	-.326*	.0972	.002	-.555	-.098
	Tamhane	Insomnia	None	-.769*	.0663	.000	-.928	-.609
			Sleep Apnea	-.442*	.1189	.001	-.731	-.154
		None	Insomnia	.769*	.0663	.000	.609	.928
			Sleep Apnea	.326*	.1210	.024	.033	.620
		Sleep Apnea	Insomnia	.442*	.1189	.001	.154	.731
			None	-.326*	.1210	.024	-.620	-.033
QualityofSleep	Tukey HSD	Insomnia	None	-1.09*	.148	.000	-1.44	-.74
			Sleep Apnea	-.67*	.180	.001	-1.10	-.25
		None	Insomnia	1.09*	.148	.000	.74	1.44
			Sleep Apnea	.42*	.148	.013	.07	.77
		Sleep Apnea	Insomnia	.67*	.180	.001	.25	1.10
			None	-.42*	.148	.013	-.77	-.07
	Tamhane	Insomnia	None	-1.09*	.113	.000	-1.37	-.82
			Sleep Apnea	-.67*	.208	.005	-1.18	-.17
		None	Insomnia	1.09*	.113	.000	.82	1.37
			Sleep Apnea	.42	.198	.104	-.06	.90
		Sleep Apnea	Insomnia	.67*	.208	.005	.17	1.18
			None	-.42	.198	.104	-.90	.06
StressLevel	Tukey HSD	Insomnia	None	.76*	.232	.003	.21	1.30
			Sleep Apnea	.20	.281	.749	-.46	.86
		None	Insomnia	-.76*	.232	.003	-1.30	-.21
			Sleep Apnea	-.55*	.231	.045	-1.10	-.01
		Sleep Apnea	Insomnia	-.20	.281	.749	-.86	.46
			None	.55*	.231	.045	.01	1.10
	Tamhane	Insomnia	None	.76*	.198	.001	.28	1.24
			Sleep Apnea	.20	.312	.887	-.55	.96
		None	Insomnia	-.76*	.198	.001	-1.24	-.28
			Sleep Apnea	-.55	.285	.157	-1.24	.14
		Sleep Apnea	Insomnia	-.20	.312	.887	-.96	.55
			None	.55	.285	.157	-.14	1.24

Based on observed means.

The error term is Mean Square(Error) = 3,057.

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

Yukarıdaki tabloda gruplar arası karşılaştırmaların özeti verilmiş ve anlamlı farklılıklar aşağıda Tamhane testi değerlerine göre yorumlanmıştır.

Örneğin uyku süresinin “Insomnia”, “None” ve “Sleep Apnea” üzerinde anlamlı bir farklılığı olduğunu ($p < 0.05$) söylenebilir.

Diğer bakılması gereken yer yine aynı tabloda bulunan “Mean Difference” değerleri olacaktır. Bu değerler üzerinde bulunan yıldızlar sonuçların anlamlı olduğunu verir. Anlamlı farklılığı olan her bir grup için aşağıda yorumlama yapılmıştır.

Uyku süresi için:

- “None” grubu, “Insomnia” ve “Sleep Apnea” gruplarına göre uyku süresini daha fazla artırıyor.

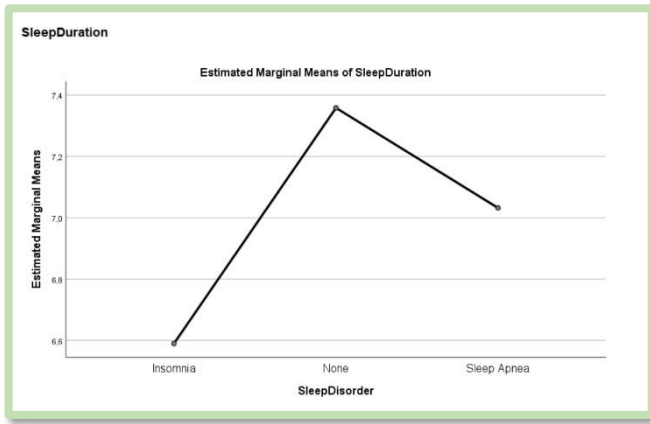
Uyku kalitesi için:

- “Insomnia” grubu, “Sleep Apnea” grubuna göre uyku kalitesini daha az düşürüyor.
- “None” grubu, “Insomnia” grubuna göre uyku kalitesini daha fazla artırıyor.

Stres seviyesi için tek anlamlı farkın “None” ile “Insomnia” grubu arasında olduğu görülmektedir.

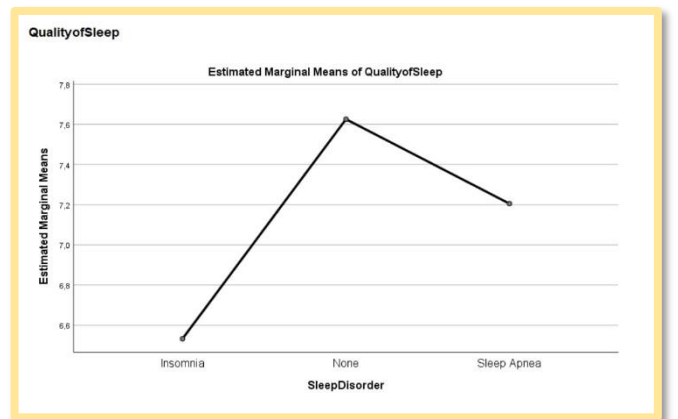
- “Insomnia” grubu, “None” grubuna göre stres seviyesini artırmaktadır.

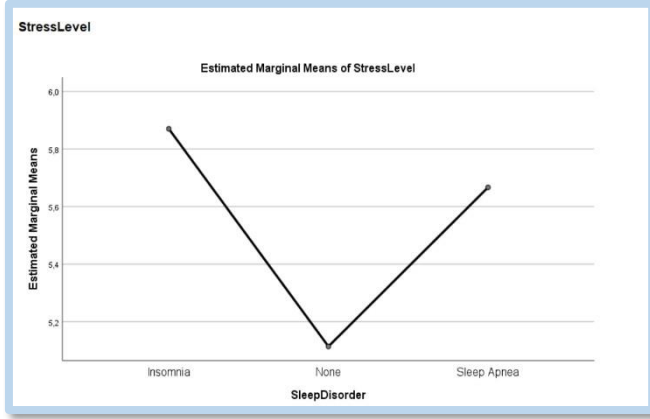
Aşağıda Manova uygulanmış gruplara ilişkin grafikler bulunmaktadır.



İlk grafiğe bakıldığında uyku süresini en çok düşüren rahatsızlığın “Insomnia”, sonrasında ise “Sleep Apnea” olduğu söylenebilir. Herhangi bir uyku problemi olmayanların uyku süresi de pozitif şekilde etkilenmektedir.

Burada ise uyku kalitesi ele alınmıştır. Benzer şekilde “Insomnia” uyku kalitesini negatif yönde en çok etkileyen rahatsızlıktır. Hemen ardından “Sleep Apnea” ‘nin de olumsuz şekilde etkilediği söylenebilir. Beklenildiği üzere hiçbir rahatsızlığı olmayan kişilerin uyku kalitesinin de pozitif şekilde arttığı görülür.





Son olarak soldaki tabloda stres seviyesi ile ilgili gruplar kıyaslanmaktadır. Uykusuzluk konusunda hiçbir rahatsızlığı olmayan bireylerin stres seviyesi de düşük çıkmıştır. Buna zıt olarak, başta “Insomnia” olmak üzere “Sleep Apnea” rahatsızlıklarına sahip kişilerin stres seviyesinin de arttığı görülür.

FAKTÖR ANALİZİ

Bu bölümde faktör analizi yapmak için öncelikle aşağıda bulunan “KMO and Bartlett’s Test” tablosunun incelenmesi gerekmektedir.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,592
Bartlett's Test of Sphericity	Approx. Chi-Square	2117,534
	df	15
	Sig.	,000

Tablo değerine bakıldığında KMO değerinin 0.592 olduğu görülmektedir. Dolayısıyla örneklem büyüklüğünün faktör analizi için yeterli olduğu söylenebilir.

Hemen altında bulunan “Bartlett’s Test of Sphericity” değerini yorumlayabilmek için hipotez kurmamız gerekmektedir.

H_0 : Korelasyon matrisi birim matristir.

H_1 : Korelasyon matrisi birim matris değildir.

Kurulan hipotez için tabloda bulunan p-value değeri 0.05’ten küçük olduğundan H_0 reddedilir. %95 güven düzeyinde korelasyon matrisinin birim matris olmadığı söylenebilir.

Inverse of Correlation Matrix						
	SleepDuration	PhysicalActivityLevel	HeartRate	DailySteps	QualityofSleep	StressLevel
SleepDuration	5,053	-1,089	-,064	1,087	-4,199	,126
PhysicalActivityLevel	-1,089	4,998	-2,616	-4,338	,039	1,886
HeartRate	-,064	-2,616	3,372	2,496	,873	-2,082
DailySteps	1,087	-4,338	2,496	5,069	-1,333	-3,085
QualityofSleep	-4,199	,039	,873	-1,333	10,082	5,321
StressLevel	,126	1,886	-2,082	-3,085	5,321	7,921

“Inverse Of Correlation Matrix” tablosunda köşegen elemanlarının incelenmesi gerekir. Köşegen elemanları arasında 5’ten büyük bir değer olması çoklu bağlantı sorununa işaret etmektedir. Tablo incelendiğinde bu durum “QualityofSleep” ve “StressLevel” değişkenlerinde karşımıza çıkmaktadır.

Anti-image Matrices							
		SleepDuration	PhysicalActivityLevel	HeartRate	DailySteps	QualityofSleep	StressLevel
Anti-image Covariance	SleepDuration	,198	-,043	-,004	,042	-,082	,003
	PhysicalActivityLevel	-,043	,200	-,155	-,171	,001	,048
	HeartRate	-,004	-,155	,297	,146	,026	-,078
	DailySteps	,042	-,171	,146	,197	-,026	-,077
	QualityofSleep	-,082	,001	,026	-,026	,099	,067
	StressLevel	,003	,048	-,078	-,077	,067	,126
Anti-image Correlation	SleepDuration	,799 ^a	-,217	-,016	,215	-,588	,020
	PhysicalActivityLevel	-,217	,352 ^a	-,637	-,862	,006	,300
	HeartRate	-,016	-,637	,551 ^a	,604	,150	-,403
	DailySteps	,215	-,862	,604	,308 ^a	-,187	-,487
	QualityofSleep	-,588	,006	,150	-,187	,731 ^a	,595
	StressLevel	,020	,300	-,403	-,487	,595	,698 ^a

a. Measures of Sampling Adequacy(MSA)

“Anti-image Matrices” tablosunda MSA değerinin her değişken için 0.5 değerinden büyük olması istenmektedir. Köşegen elemanları incelendiğinde “PhysicalActivityLevel” ve “DailySteps” değişkenlerinin istenen değerden düşük çıktığı görülmüş, değerlerin 0.5’ten küçük olduğu varsayılarak analize devam edilmiştir.

“Communalities” tablosunda bulunan “Extraction” değerleri, değişkenlerin açıklanma oranlarını vermektedir.

Communalities

	Initial	Extraction
SleepDuration	1,000	,828
PhysicalActivityLevel	1,000	,910
HeartRate	1,000	,601
DailySteps	1,000	,868
QualityofSleep	1,000	,941
StressLevel	1,000	,910

Extraction Method: Principal Component Analysis.

Yandaki tabloda görüldüğü üzere tüm değişkenlerin açıklama oranları %50'den büyüktür.

Analiz başında “Communalities” tablosunda görülen değişkenlerin yanı sıra “Age” değişkeni de incelenmiştir. Ancak sözü edilen değişkenin değeri “0.325” geldiği ve dolayısıyla açıklama oranı %50'den düşük olduğu için bu değişken faktör analizinden çıkartılmıştır.

“KMO and Bartlett’s Test” tablosu da sözü edilen değişkenin çıkartılıp, yeniden yapılan analiz sonuçlarına göre yorumlanmıştır.

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	3,252	54,202	54,202	3,252	54,202	54,202	3,250	54,159	54,159
2	1,805	30,088	84,291	1,805	30,088	84,291	1,808	30,132	84,291
3	,627	10,443	94,733						
4	,165	2,744	97,477						
5	,093	1,557	99,034						
6	,058	,966	100,000						

Extraction Method: Principal Component Analysis.

“Total Variance Explained” tablosuna bakıldığında “% of Variance” için değerler, analizde kullanılan faktörlerin veya temel bileşenlerin toplam varyansın yüzde kaçını açıkladığını göstermektedir. “Cumulative %” değerleri ise analizin belirli bir sayıdaki ilk faktör veya temel bileşeni kullanarak toplam varyansın yüzde kaçını açıkladığını gösterir.

Veri setinde toplamda 6 değişken varken SPSS’in bunu 2 değişkene indirgemiş olduğu görülmektedir. Birinci değişkenin tek başına veriyi açıklama oranının %54, ikinci değişkenin veriyi tek başına açıklama oranının ise %30 olduğu söylenebilir. Toplamda %84 ile bu iki faktörün veriyi açıklama oranı yeterlidir.

Component Matrix^a

	Component	
	1	2
QualityofSleep	,967	
StressLevel	-,945	
SleepDuration	,907	
HeartRate	-,764	
PhysicalActivityLevel		,946
DailySteps		,931

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Soldaki tablo, oluşturulan iki ayrı faktör için değişkenlerin hangisine atandığını göstermektedir.

Birinci faktöre uyku kalitesi, stres seviyesi, uyku süresi ve kalp atış hızını gösteren değişkenler yüklenmişken; ikinci faktöre fiziksel aktivite ve günlük atılan adım değişkenleri yüklenmiştir.

Tabloda görüldüğü üzere çapraz yüklenme söz konusu değildir.

Tablodan hareketle faktör

denklemleri aşağıdaki gibi oluşturulur;

- $QualityofSleep = 0.96 F_1 + \Sigma$
- $StressLevel = -0.94 F_1 + \Sigma$
- $SleepDuration = 0.90 F_1 + \Sigma$
- $HeartRate = -0.76 F_1 + \Sigma$
- $PhysicalActivityLevel = 0.94 F_2 + \Sigma$
- $DailySteps = 0.93 F_2 + \Sigma$

DISKRIMINANT ANALIZI

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
SleepDisorder	String	11	0		{Insomnia, 3...	None	11	Left	Nominal	Input
Disorders2	Numeric	11	0		{1, None}...	None	11	Right	Nominal	Input

SleepDisorder	Disorders2	Age	SleepDuration	QualityofSleep	PhysicalActivityLevel	StressLevel	HeartRate	DailySteps	var
1	None	27	6.1	6	42	6	77	4200	
1	None	28	6.2	6	60	8	75	10000	
1	None	28	6.2	6	60	8	75	10000	
2	Sleep Apnea	28	5.9	4	30	8	85	3000	
2	Sleep Apnea	28	5.9	4	30	8	85	3000	
3	Insomnia	28	5.9	4	30	8	85	3000	
3	Insomnia	29	6.3	6	40	7	82	3500	
1	None	29	7.8	7	75	6	70	8000	
1	None	29	7.8	7	75	6	70	8000	
1	None	29	7.8	7	75	6	70	8000	
1	None	29	6.1	6	30	8	70	8000	
1	None	29	7.8	7	75	6	70	8000	
1	None	29	6.1	6	30	8	70	8000	
1	None	29	6.1	6	30	8	70	8000	
1	None	29	6.0	6	30	8	70	8000	
1	None	29	6.0	6	30	8	70	8000	
1	None	29	6.0	6	30	8	70	8000	
1	None	29	6.0	6	30	8	70	8000	
2	Sleep Apnea	29	6.5	5	40	7	80	4000	
2	Sleep Apnea	29	6.0	6	30	8	70	8000	
3	Insomnia	29	6.5	5	40	7	80	4000	
1	None	30	7.6	7	75	6	70	8000	
1	None	30	7.7	7	75	6	70	8000	
1	None	30	7.7	7	75	6	70	8000	
1	None	30	7.7	7	75	6	70	8000	
1	None	30	7.7	7	75	6	70	8000	
1	None	30	7.7	7	75	6	70	8000	
1	None	30	7.8	7	75	6	70	8000	
1	None	30	7.9	7	75	6	70	8000	

Diskriminant analizini gerçekleştirmeden önce “SleepDisorder” değişkeni, orjinal veri setinde “string” formatında olduğu için farklı bir sayfa açılıp “numeric” forma dönüştürülmüş ve “Disorders2” olarak isimlendirilmiştir.

(1:None 2:Sleep Apnea 3:Insomnia)

Test Results

Box's M	1054,588
F	Approx. 18,175
	df1 56
	df2 140342,945
	Sig. ,000

Tests null hypothesis of equal population covariance matrices.

Bu analize başlamadan önce hipotez testini kuralım;

H₀: Tüm bağımsız değişkenlerin grupları arasındaki farkı açıklama yetenekleri aynıdır.

H₁: En az bir bağımsız değişkenin grupları arasındaki farkı açıklama yeteneği diğerlerinden farklıdır.

p-value değeri 0.05'ten küçük olduğu için H₀ hipotezi reddedilmiştir.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	1,738 ^a	75,8	75,8	,797
2	,554 ^a	24,2	100,0	,597

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

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,235	532,889	14	,000
2	,643	162,236	6	,000

Yukarıda bulunan iki ayrı tabloya birlikte bakmak daha mantıklı olacaktır. “Eigenvalues” değerleri her bir değişken için veri setinin açıklanan varyansını verirken, “Wilks’ Lambda” ise açıklanamayan varyansı vermektedir.

İlk tabloda bulunan “Canonical Correlation” değerlerinin kareleri alındığında, birinci fonksiyonun veri seti varyansını açıklama oranı %63.5 iken ikinci fonksiyon için oran %35.6 gelmektedir. Aynı zamanda özdeğerimiz 0.4’ten büyük olduğu için verinin anlamlı olduğu söylenebilir.

İkinci tabloya bakılacak olursa, “Wilks’ Lambda” değerlerinin karesi alındığında ilk fonksiyon için açıklanamayan varyans oranı %5, ikinci fonksiyon için %41.3 gelmektedir.

Structure Matrix		
	Function	
	1	2
Age	-,409 [*]	,220
HeartRate	-,301 [*]	,206
StressLevel	-,133 [*]	-,085
PhysicalActivityLevel	-,133	,635 [*]
DailySteps	-,037	,514 [*]
SleepDuration	,244	,342 [*]
QualityofSleep	,219	,336 [*]

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

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

Soldaki tabloda bir değişkenin katsayısının mutlak büyüklüğü ile modele katkısı doğru orantılıdır.

Tablo incelendiğinde birinci fonksiyona en büyük katkı sağlayan değişkenin 0.41 değeri ile “Age”, ikinci fonksiyona en büyük katkı sağlayan değişkenin 0.63 ile “PhysicalActivityLevel” olduğu görülmektedir.

Prior Probabilities for Groups

Disorders2	Prior	Cases Used in Analysis	
		Unweighted	Weighted
None	,586	219	219,000
Sleep Apnea	,211	79	79,000
Insomnia	,203	76	76,000
Total	1,000	374	374,000

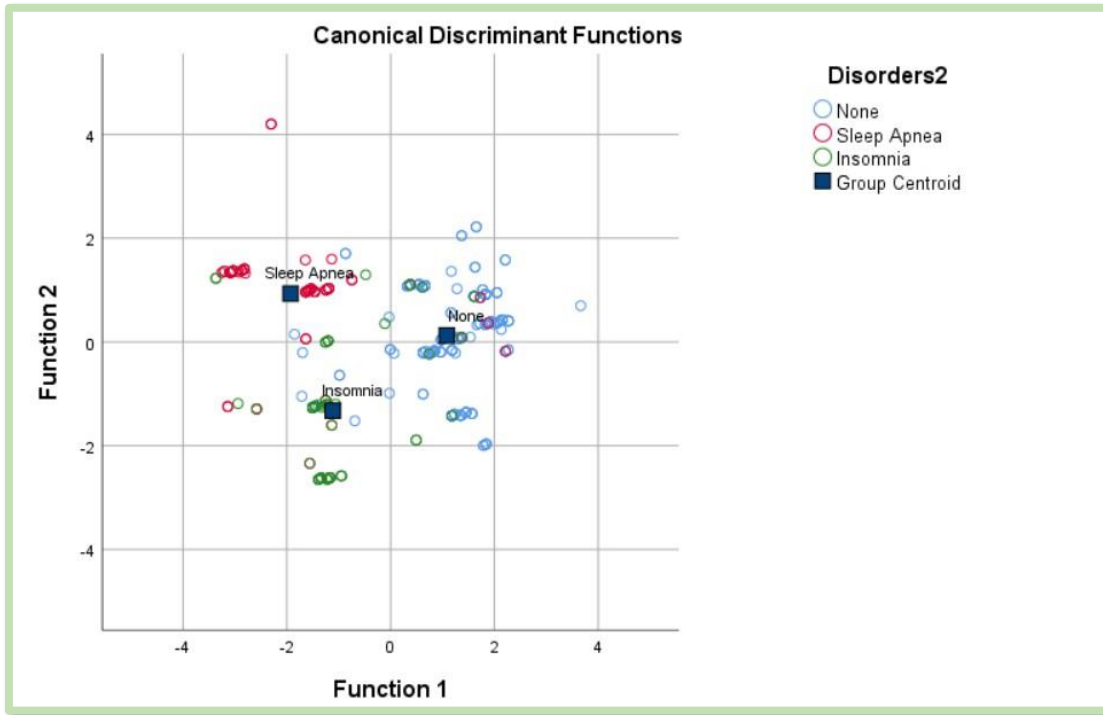
“Prior Probabilities for Groups” tablosunda değışkende bulunan kategoriler incelendiğinde öncelikli grubun %58.6 ile “None” grubu olduğu söylenebilir. Sonrasında sırasıyla “Sleep Apnea” ve “Insomnia” grupları gelmektedir.

Classification Results^{a,c}

		Disorders2	Predicted Group Membership			Total
			None	Sleep Apnea	Insomnia	
Original	Count	None	202	9	8	219
		Sleep Apnea	5	66	8	79
		Insomnia	11	8	57	76
	%	None	92,2	4,1	3,7	100,0
		Sleep Apnea	6,3	83,5	10,1	100,0
		Insomnia	14,5	10,5	75,0	100,0
Cross-validated ^b	Count	None	202	9	8	219
		Sleep Apnea	5	66	8	79
		Insomnia	11	8	57	76
	%	None	92,2	4,1	3,7	100,0
		Sleep Apnea	6,3	83,5	10,1	100,0
		Insomnia	14,5	10,5	75,0	100,0

Yukarıdaki tabloya ait yorumlar aşağıdaki gibi özetlenebilir.

- Orijinal gruplandırılmış vakaların %86.9'u doğru şekilde sınıflandırılmıştır.
- Çapraz doğrulamalı grup vakalarının %86.9'u doğru şekilde sınıflandırılmıştır.
- Genel olarak %86.9 doğruluk oranı, modelin veri setinde ve çapraz doğrulama işlemi sırasında doğru sınıflandırma oranının yüksek olduğunu göstermektedir.



Diskriminant analizinde incelenecek son görsel olan “Canonical Discriminant Functions” grafiği yukarıda verilmiştir. Bu grafikten hareketle oluşturulan lineer model aşağıdaki gibidir.

$$CDF = a_1 \text{ SleepApnea} + a_2 \text{ None} + a_3 \text{ Insomnia} + b_1 \quad (b_1 : \text{sabit terim})$$

LOJİSTİK REGRESYON

Model Fitting Information						
Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	603,677	611,526	599,677			
Final	269,556	316,647	245,556	354,122	10	,000

“Model Fitting Information” tablosunda bulunan “-2 Log Likelihood” “Final” değeri ne kadar küçükse test o kadar anlamlıdır.

Pseudo R-Square

Cox and Snell	,612
Nagelkerke	,716
McFadden	,490

R^2 değeri büyük olan daha çok tercih edilmektedir. Tabloda 0.716 ile “Nagelkerke” tercih edilmiştir.

Likelihood Ratio Tests

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	269,556	316,647	245,556 ^a	,000	0	.
HeartRate	319,410	358,653	299,410	53,855	2	,000
SleepDuration	320,527	359,770	300,527	54,972	2	,000
BMIcategory	495,844	519,390	483,844	238,289	6	,000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Yukarıdaki tablo, ilgili değişkenlerin çıkarıldıkları zaman analize olumsuz etkiye bulunup bulunmadığını göstermektedir. Bunun için p-value değerine veya “-2 Log Likelihood Of Reduced Model” “Intercept” değeri ile diğer değerler karşılaştırılabilir.

“-2 Log Likelihood Of Reduced Model” için bakılacak olursa, “Intercept” değerine göre değişkenlerin değerleri çıkarıldığı takdirde hep büyük geldiğinden değişkenlerin anlamlı olduğu söylenebilir.

Parameter Estimates								
SleepDisorder ^a		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B) Lower Bound Upper Bound
Insomnia	Intercept	54,954	9,906	30,773	1	,000		
	HeartRate	-,536	,105	26,279	1	,000	,585	,477 ,718
	SleepDuration	-2,473	,424	34,095	1	,000	,084	,037 ,193
	[BMICategory=Normal]	,538	,670	,645	1	,422	1,713	,461 6,368
	[BMICategory=Normal Weight]	1,843	1,146	2,583	1	,108	6,314	,667 59,729
	[BMICategory=Obese]	6,491	1,541	17,734	1	,000	659,111	32,132 13519,915
	[BMICategory=Overweight]	0 ^b	.	.	0	.	.	.
None	Intercept	34,750	10,112	11,809	1	,001		
	HeartRate	-,379	,107	12,481	1	,000	,685	,555 ,845
	SleepDuration	-1,283	,427	9,017	1	,003	,277	,120 ,640
	[BMICategory=Normal]	4,645	,564	67,777	1	,000	104,080	34,442 314,517
	[BMICategory=Normal Weight]	4,314	,937	21,214	1	,000	74,735	11,920 468,561
	[BMICategory=Obese]	-14,832	,000	.	1	.	3,619E-7	3,619E-7 3,619E-7
	[BMICategory=Overweight]	0 ^b	.	.	0	.	.	.

a. The reference category is: Sleep Apnea.

b. This parameter is set to zero because it is redundant.

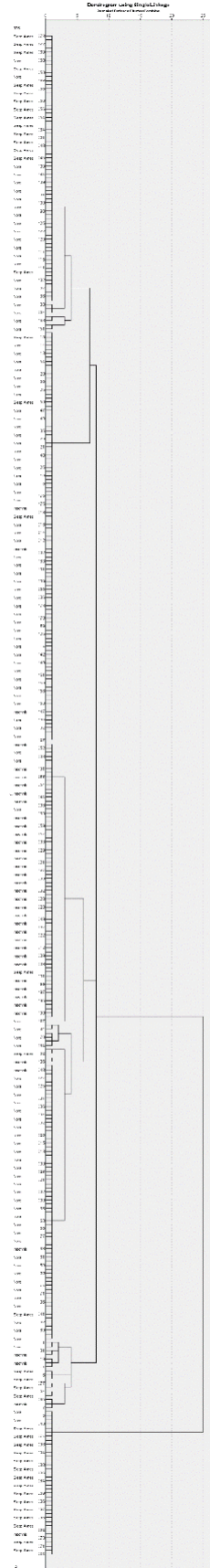
“Parameter Estimates” tablosunda referans kategorisi “Sleep Apnea” olduğu için yorumlama da ona göre şekillenmelidir.

Tabloya bakıldığında, “Insomnia” için vücut kitle indeksi normal olan bireylerin ilgili değişken ile anlamlı bir ilişkiye sahip olmadığı söylenebilir. Onun dışındakiler için “ $p < 0.05$ ” olduğundan ilişkiler anlamlıdır.

Yorumlanacak olursa:

- “HeartRate” kategorisindeki bir birimlik artış None ile karşılaştırılmak istendiğinde, kişide uyku bozukluğu açısından “Sleep Apnea” rahatsızlığı olasılığını %38 artırmaktadır.
- Insomnia ile karşılaştırıldığında, “HeartRate”te bir birimlik artış kişide uyku bozukluğunda “Sleep Apnea” rahatsızlığı olasılığını %53,6 artırmaktadır.
- Insomnia için, bireyin vücut kitle indeksinin obez derecesinde olması uyku bozukluğu açısından “Insomnia” rahatsızlığına yakalanma olasılığını %154 artırmaktadır.

KÜMELEME



Yandaki grafiğe bakıldığında veri setinin 3 kümeye ayrıldığı görülmektedir. Dolayısıyla k değeri 3 alınarak analize devam edilecektir.

Final Cluster Centers			
	Cluster		
	1	2	3
Disorders2	2	2	1
SleepDuration	7,3	6,2	7,2
PhysicalActivityLevel	63	88	38
HeartRate	69	74	72
DailySteps	7137	10000	4793

“Final Cluster Center” tablosu küme ortalamalarını vermektedir ve aşağıdaki gibi yorumlanabilir.

- Fiziksel aktivite açısından en yüksek ortalamaya sahip olan “88” ile 2. kümedir. Sonrasında sırasıyla 1. ve 3. küme gelmektedir.
- Kalp atış hızı açısından en yüksek olan yine 2. küme olup, sonrasında 3. küme gelir. En düşük küme ise 1. kümedir.
- Günlük atılan adım açısından en yüksek ortalamaya sahip olan 2. küme, sonrasında ise 1. kümedir. En düşük küme 3. kümedir.

Distances between Final Cluster Centers			
Cluster	1	2	3
1		2863,490	2344,132
2	2863,490		5207,614
3	2344,132	5207,614	

“Distances between Final Cluster Centers” tablosu kümelerin birbirine olan uzaklığını vermektedir. Yukarıda yapılmış olan yorumlar ile birlikte bu tabloda da görüldüğü üzere 2. ve 3. kümenin birbirine en uzak iki küme olduğu söylenebilir.

ANOVA						
	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Disorders2	3,937	2	,631	371	6,236	,002
SleepDuration	18,001	2	,539	371	33,369	,000
PhysicalActivityLevel	38802,586	2	227,083	371	170,874	,000
HeartRate	577,551	2	14,083	371	41,012	,000
DailySteps	389437501,0	2	532368,935	371	731,518	,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 tablosunda ortalama deęerlerin anlamlı farklılıklar göstermesi beklenir. Bu duruma uygun hipotez testi ve yorumu ařaęıda verilmiřtir.

H₀: Kme merkezleri arasında bir fark yoktur.

H₁: En az bir kme merkezi dięerlerinden farklıdır.

Yukarıda ele alınan hipotezin p-value deęerlerine bakarak yorumlanması gerekmektedir.

Hepsi iin “ $p < 0.05$ ” durumu sz konusu olduęundan H_0 reddedilir. Kmeleme analizinin bařarılı ve en az iki kme arasında anlamlı bir fark olduęu sylenebilir.