



BIYOİSTATİSTİK

Dr. Öğr. Üyesi Kürşad Nuri BAYDİLİ

Sağlık Bilimleri Üniversitesi

Hamidiye Tıp Fakültesi

Biyoistatistik ve Tıbbi Bilişim Ana Bilim Dalı

İS-TA-TİS-TİK

İstatistik sözcüğü ilk olarak Alman devlet verilerinin sunumunda 1749 yılında kullanılmıştır.

"Statistik"

political arithmetic-siyasi aritmetik-Devlet bilimi

İstatistik

- İstatistiğin temelleri ise çok daha eskilere MÖ 5. yüzyıla dayanmaktadır. Hippias, ilk olimpiyatların gerçekleştirildiği tarihi tespit edebilmek için kralların ortalama tahtta kalış sürelerini göz önüne alınarak hesaplamalar gerçekleştirmiştir.



«Aritmetik ortalama»

İstatistik

- MS 2 Yılında Çin'de, bilinen ilk nüfus sayımı işlemi gerçekleştirilmiştir.

«Tam sayım»

- MS 9. Yüzyılda ise El-Kındi, Jül Sezar tarafından bulunan ve bahsi geçen dönemde güncelliğini koruyan tek alfabeli şifreleme yerine harflerin sıklıklarını kullanarak yeni bir kriptografi tekniği bulmuştur.

«Frekans-sıklık analizi»

İSTATİSTİK

- 12. Yüzyılda İngiliz Darphanesi madeni paralardaki alařım oranlarını kontrol etmek için basılan paralardan belirli periyotlarda, belirli sayılarda alarak kontrol etmişlerdir.

«Örnekleme»



- İstatistiğin matematiksel temelleri ise 1654 yılında Pierre Fermat ve Blaise Pascal arasındaki mektuplaşmalara dayandığı kabul edilmektedir.
- 1713 Yılında Jakob Bernoulli tarafından yayımlanan Şans Doktrini (Doctrine of Chances) eseri ise istatistiği matematiğin bir dalı olarak ele almıştır.

«Şans, hazır olan kafalara yardım eder»
'Louis Pasteur'



İstatistik

- 20. Yüzyılda meydana gelen gerek pandemi ve gerekse savaş gibi toplum parametrelerini doğrudan etkileyen olaylar neticesinde sağlık, ekonomi ve sosyal alanlarda daha titiz araçlara ihtiyaç duyulması istatistiksel uygulamalarda ilerlemeyi zorunlu kılmıştır.



İSTATİSTİK

- İstatistik sözcüğünün günümüzdeki anlamına en yakın haliyle kullanımı 19. Yüzyılın başlarındadır. Sır John Sinclair tarafından İngilizceye aktarılarak kullanılan sözcük.
- ¹İstatistik: Bir sonuca varmak üzere, olayları yöntemli bir şekilde toplayıp sayısal olarak gösterme işi ve bilimidir.
- ²İstatistik: Etkin kararlar vermek amacıyla verileri toplama, sınıflama, sunma, çözümleme ve yorumlama biçimi olarak da tanımlanabilir.



İSTATİSTİK

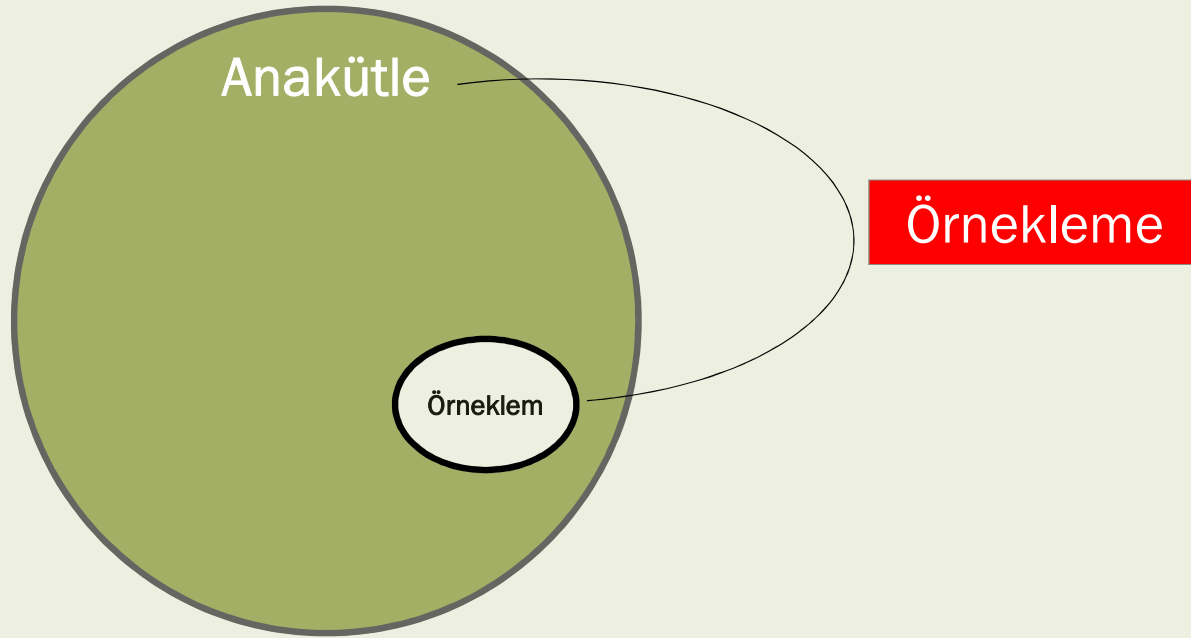
Bilim Dalları

- Matematiksel istatistik: İstatistik teorisinin matematiksel temellerini kuran, yeni kuramsal yaklaşımlarla teknikler üreten bir istatistik bilim dalıdır.
- Uygulamalı istatistik: Matematiksel istatistiğin geliştirdiği teorileri ve teknikleri çeşitli alanlarda uygulayan, bu alanlardaki işleyişlerini kontrol eden ve bu tekniklerin uygulama alanlarına özgü uyarlamalar yapan, yeni teknikler geliştiren bir istatistiksel bilim dalıdır.
- Biyoistatistik: Matematik ve istatistik tekniklerinin tıp ve sağlık bilimlerinde uygulamalarını içeren, bu alana özgü uyarlamalar yapan, yeni teknikler üreten bir bilim dalıdır.

ANA KÜTLE (EVREN)

- Hakkında bilgi edinilmek istenen topluluğun tamamına **ana kütle (evren)** denir. Ana kütle birimlerden oluşur. Ana kütlelerin yığın olaydan ve aynı cins birimlerden oluşması gerekmektedir.
- Ana kütledeki tüm birimlere ulaşılarak istenen bilginin elde edilmesi durumunda yapılan işleme **tamsayım** adı verilir.

ÖRNEKLEM VE ÖRNEKLEME

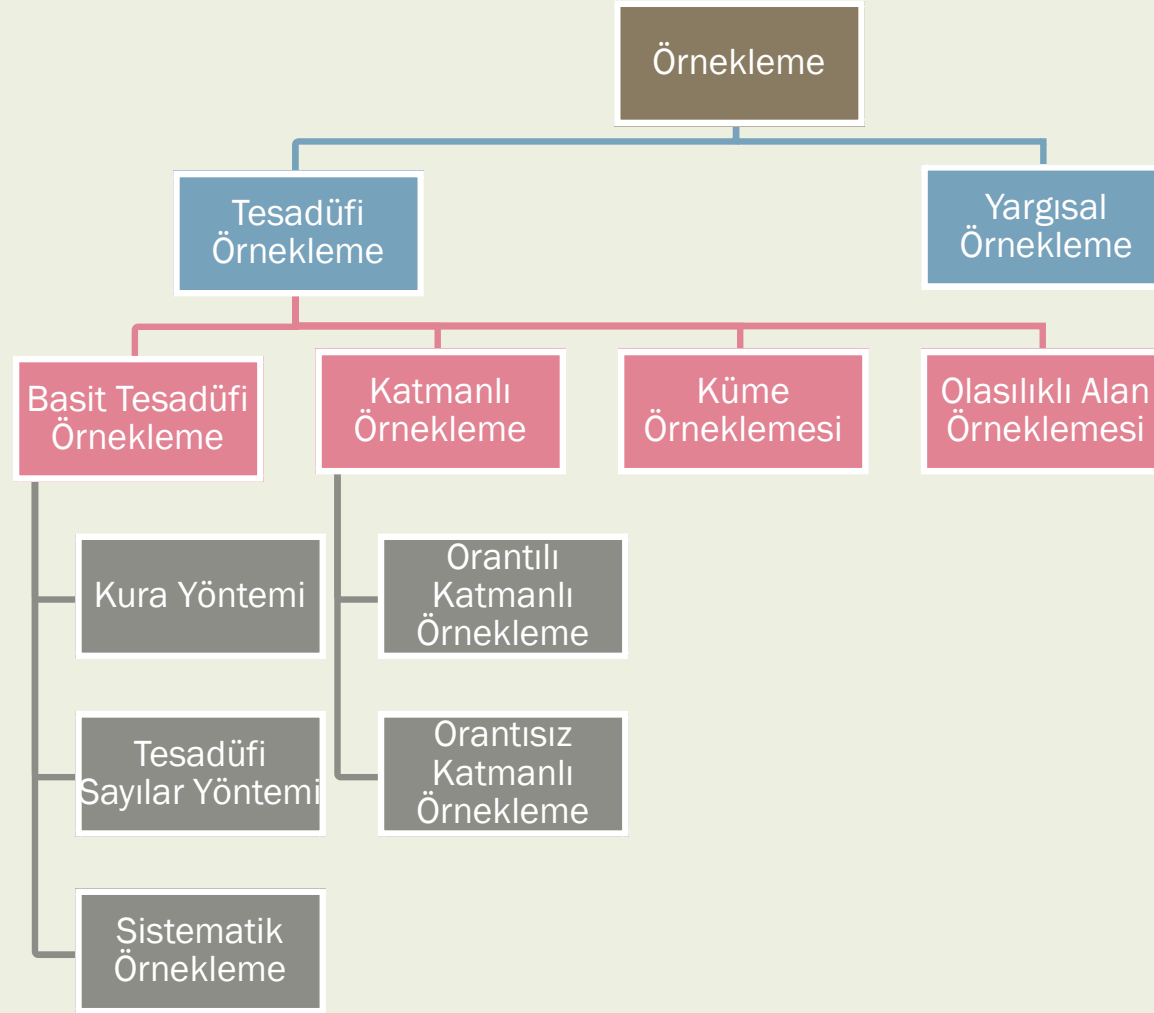


GÖZLEM

- Ana kütle ya da örnekleme de yer alan her birime gözlem denir.
- Ana kütledeki Gözlem Sayısı: N
- Örnekleme deki Gözlem Sayısı: n

ÖRNEKLEM BÜYÜKLÜĞÜ HESAPLAMALARI & *Power Analizi*

- Power analizi; araştırmanızın ana hipotezinin kabul ya da reddedilmesini sağlayacak olan istatistiksel hipotez testinin gücünü hesaplar.
- Örneklem büyüklüğü hesaplamalarının yapılabilmesi için;
 - *Araştırmanın ana hipoteziyle ilgili daha önceden yapılmış çalışmalardaki verilerden faydalanılabilir.*
 - *Pilot bir çalışma yapılarak analiz için gerekli değerler elde edilebilir.*
 - *Eğer konuyla ilgili daha önceden yapılmış bir çalışma yoksa ve pilot uygulama yapılamıyorsa ilgili değerler tahmin edilerek analizler gerçekleştirilebilir.*



PARAMETRE & İSTATİSTİK

- Ana kütle özelliklerinin sayısal değerlerine **parametre** denir. Araştırma ana kütle yerine örneklem üzerinde uygulanıyorsa, parametre değerleri tahmin edilir.
- Bir örneklemden elde edilen ve popülasyon parametresini tahmin etmek için kullanılan değere **istatistik** değerdir.

DEĐİŐKEN

- DeđiŐkenin kelime anlamı; DeđiŐme özelliđi gösteren, çok deđiŐen, deđiŐebilir, kararsız, deđiŐici Őeklindedir. Matematiksel tanımı ise; gözlemden gözleme farklı deđerler alabilen objelere, özelliklere ya da durumlara **deđerŐken** denir. Nicel (kantitatif) ve nitel (kalitatif) olmak üzere iki farklı deđerŐken tipi vardır.



NİCEL (KANTİTATİF) DEĞİŞKEN

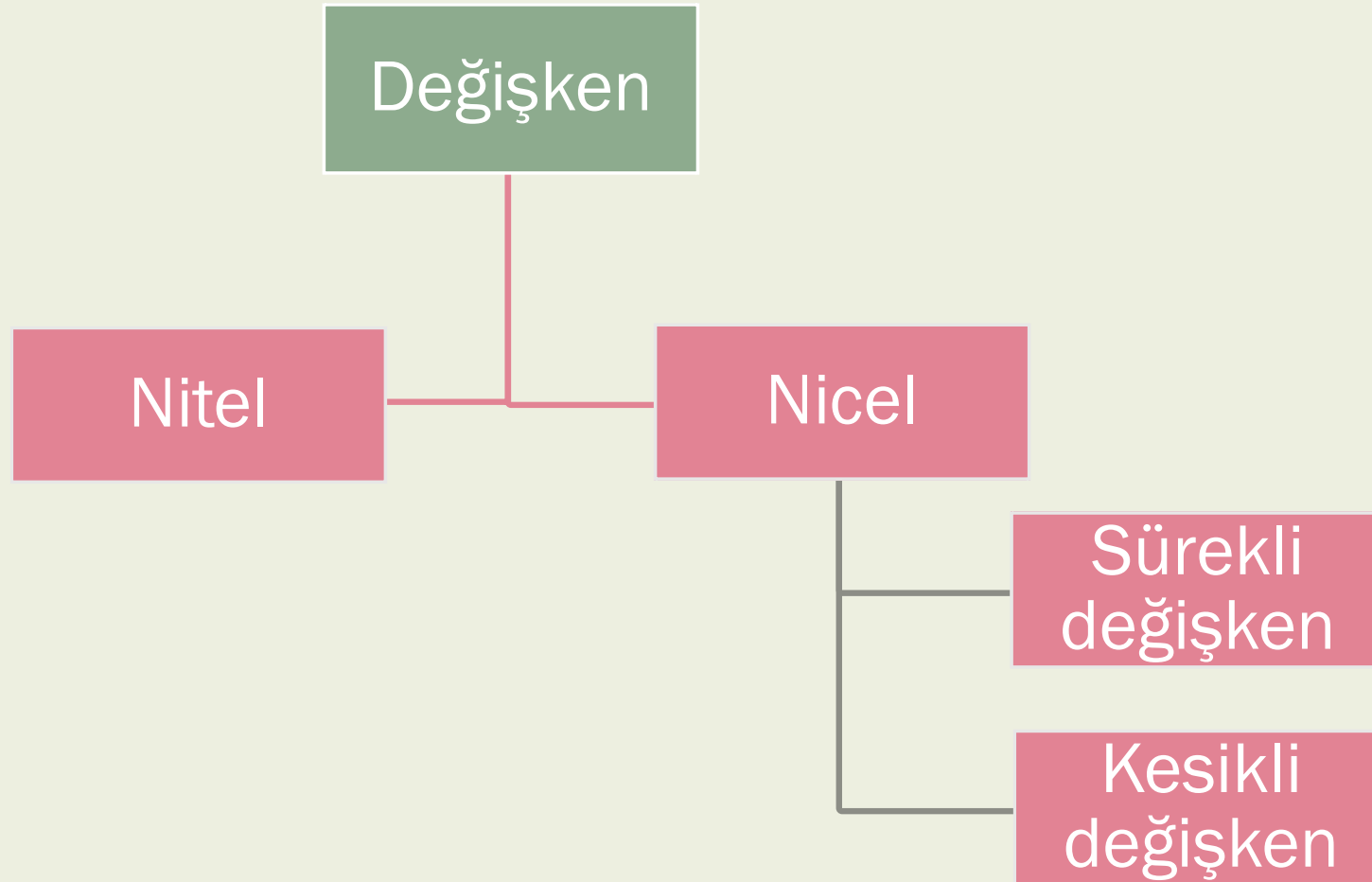
- Birimlerin ölçüm ve tartım sonucu değerleri saptanan sayısal özelliklerini belirten değişkenlerdir. Bu değişkenler değerleri, mekanik ve elektronik araçlara sayısal olarak aralıklı ölçekli yada orantılı ölçekli verileridir.
- Nicel değişkenlerin verilerine nicel veri adı verilir. Örneğin birimlerin, boy uzunluğu, vücut ağırlığı, kilosu, kan basıncı gibi özellikler nicel değişkenlerdir.





NİTEL (KALİTATİF) DEĞİŞKEN

- Nitel değişkenler; birimlerin kalite, kategorik, yada isimsel olarak belirtilebilen karakteristik özelliklerini, durumlarını ve pozisyonlarını belirten değişkenlerdir.
- Bu değişkenlerin verileri isimsel ya da sıralı ölçekle elde edilmişlerdir ve iki yada daha fazla kategoriye (alt seçenek, sınıf, grup) ayrılarak sayımla elde edilir.
- Bu değişkenlerin verilerine nitel veriler adı verilir.
- Örneğin birimlerin, cinsiyeti, kan grubu, medeni durum, göz rengi, mesleği, yerleşim yeri, tuttuğu futbol takımı (fanatikler için) gibi nitelik bildiren durumları açıklayan değişkenlerdir.



NASIL KIYASLARIZ

1

		Hastalık		Total	
		Sağlıklı	Hasta		
Cinsiyet	Erkek	Count	10	90	100
		% within Hastalık	33,3%	47,4%	45,5%
	Kadın	Count	20	100	120
		% within Hastalık	66,7%	52,6%	54,5%
Total		Count	30	190	220
		% within Hastalık	100,0%	100,0%	100,0%

2

			Hastalık		Total
			Sağlıklı	Hasta	
Cinsiyet	Erkek	Count	10	90	100
		% within Cinsiyet	10,0%	90,0%	100,0%
	Kadın	Count	20	100	120
		% within Cinsiyet	16,7%	83,3%	100,0%
Total		Count	30	190	220
		% within Cinsiyet	13,6%	86,4%	100,0%

NASIL KIYASLARIZ

- Öğrencilerin istatistik ve matematik derslerinden aldığı ders notları tablodaki gibidir.

Buna göre öğrenciler hangi derste daha başarılıdır?

Öğrenci	İstatistik	Matematik
Ali	10	5
Ayşe	15	10
Ahmet	20	15
Mehmet	25	20
Elif	30	25

	İstatistik	Matematik
Aritmetik ortalama (\bar{x})	20	15

NASIL KIYASLARIZ

- Son eklenen öğrenci ile birlikte öğrenciler hangi derste daha başarılıdır?

Öğrenci	İstatistik	Matematik
Ali	10	5
Ayşe	15	10
Ahmet	20	15
Mehmet	25	20
Elif	30	25
Kürşad	35	100

	İstatistik	Matematik
Aritmetik ortalama (\bar{x})	22,5	29,17
Ortanca (medyan- Q_2)	22,5	17,5

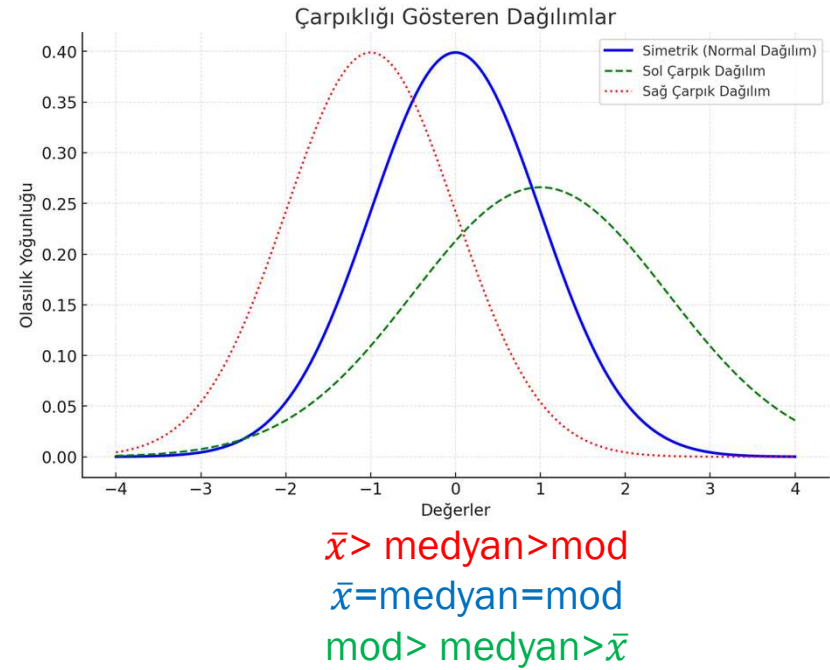
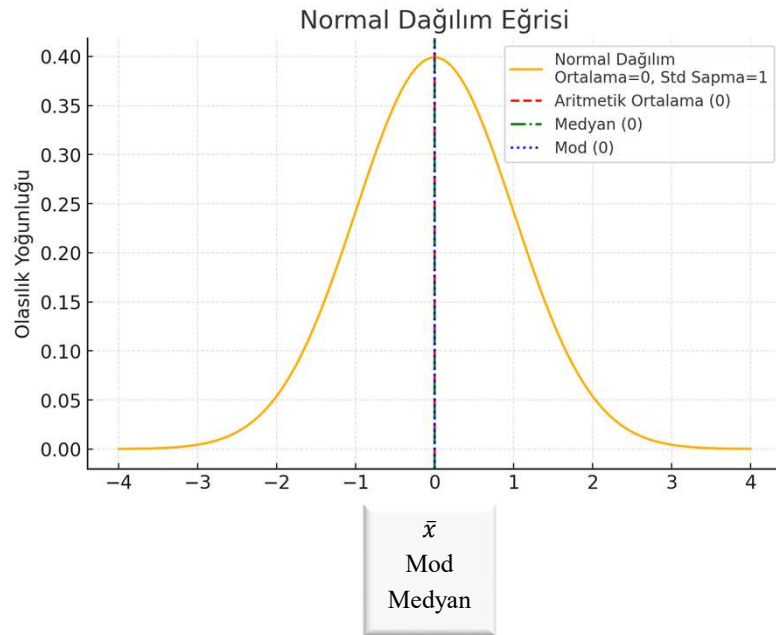
ORTALAMA ya da MEDYAN YETERLİ Mİ?

- A= (0, 25, 50, 75, 100)
- B= (48, 49, 50, 51, 52)

	A	B
Standart sapma	39,53	1,58
Varyans	1562,5	2,5
OMS	30	1,2

NORMAL DAĞILIM

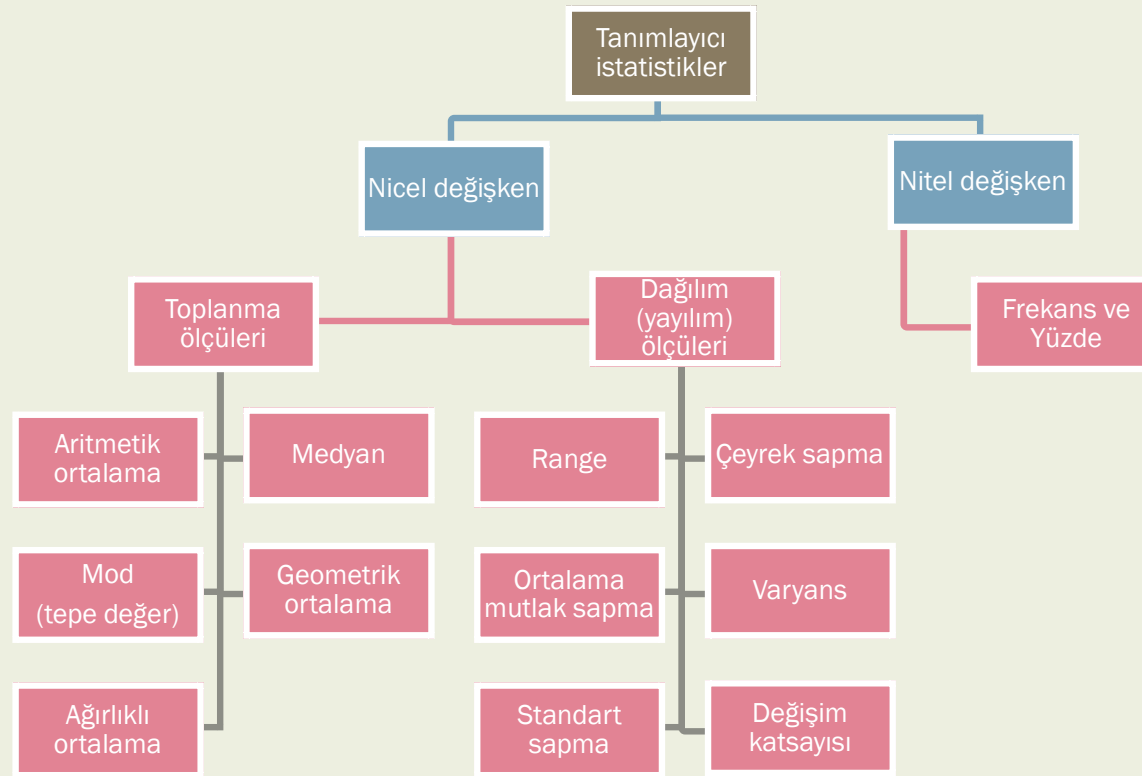
- Frekans dağılımları çan eğrisi biçiminde, simetrik görünüşe sahip olan sürekli değişkenlerin uyduğu bir dağılım türüdür.

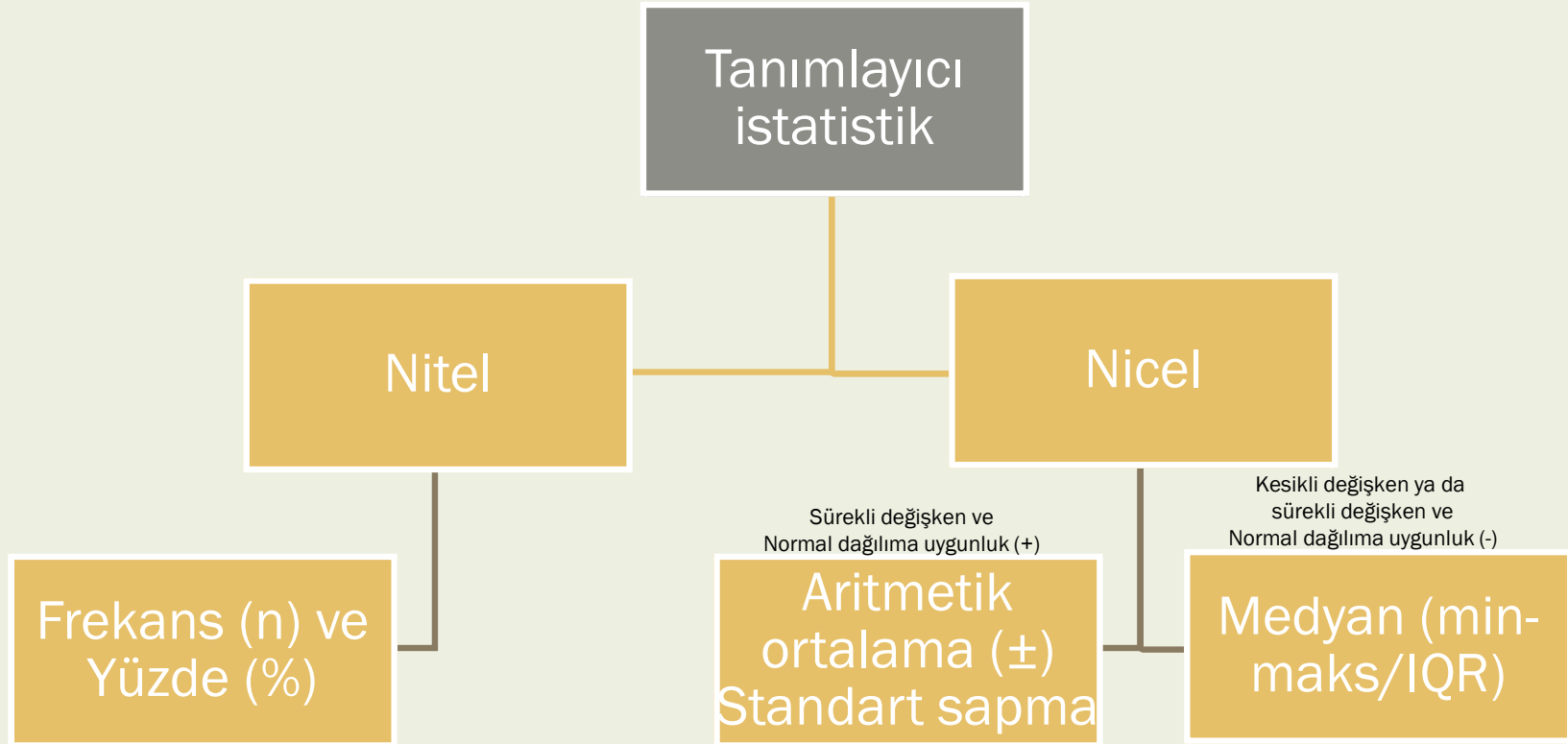


NORMAL DAĞILIM

- 68-95-99.7 Kuralı (Empirik Kural):
 - Verilerin %68'i ortalamanın ± 1 standart sapma içinde yer alır.
 - Verilerin %95'i ortalamanın ± 2 standart sapma içinde yer alır.
 - Verilerin %99.7'si ortalamanın ± 3 standart sapma içinde yer alır.

TANIMLAYICI İSTATİSTİKLER





SADECE TANIMLAYICI İSTATİSTİK YETERLİ Mİ?

➤ **Hipotez**

Kuramsal olarak varsayılan veya önceden yapılmış bir dizi gözleme dayanarak ortaya atılan, gerçekleşmesi söz konusu olan ya da olmayan bir önermedir.

➤ **Araştırma Hipotezi**

Bir araştırmada ortaya çıkarılması istenilen, beklenen sonuçların veya amaçların her birinin olumlu, olumsuz ya da soru cümleleri ile ifade edilmiş önermeleridir. Araştırmacının araştırmadan beklediği sonuçları ifade eden cümledir.

İstatistiksel Hipotez

İstatistiksel hipotezler “sıfır hipotezi” ve “alternatif hipotez” olmak üzere ikiye ayrılır.

- **Sıfır Hipotezi:** İncelenen konu ile ilgili (ilişki, farklılık, değişim) yokluğu savunan hipotezdir. H_0 sembolü ile ifade edilir.
- **Alternatif Hipotez:** İncelenen konu ile ilgili (ilişki, farklılık, değişim) varlığı savunan hipotezdir. H_1 sembolü ile ifade edilir.

*Tablo değeri < Hesap değeri veya $p < 0,05$ olması durumunda H_0 hipotezi **reddedilir.***

p değeri nedir?

"Araştırılan olayın tesadüfen meydana gelmiş olma olasılığı"

İstatistiksel Hipotez

➤ Tip I Hata (α)

H_0 hipotezi gerçekte doğru iken yanlış kabul edilmesi. Bir diğer deyişle H_0 hipotezi kabul edilmesi gerekirken reddedilerek H_1 hipotezinin kabul edilmesi.

➤ Tip II Hata (β)

H_0 hipotezi gerçekte yanlış iken doğru kabul edilmesi. Bir diğer deyişle H_0 hipotezinin reddedilmesi gerekirken kabul edilmesi durumu.

	Gerçek Durum	
Tahmin ↓	H_0 Kabul	H_1 Kabul
H_0 Kabul	-	Tip II Hata (β)
H_1 Kabul	Tip I Hata (α)	-

Güç & Güven Düzeyi

➤ Hipotez Testlerinde Güven Düzeyi ve Güç Kavramları

Hipotez testlerinde;

- Güven düzeyi $1-\alpha$,
- Güç ise $1-\beta$

Şeklinde ifade edilir. Güven düzeyi ve gücü yüksek bir karşılaştırma için α ve β değerlerinin (tip I ve tip II hata oranlarının) minimize edilmesi gerekmektedir. α ve β değerlerini minimum düzeyde tutmanın en kolay yolu örneklem büyüklüğünü artırmaktır.

$$p < \alpha$$

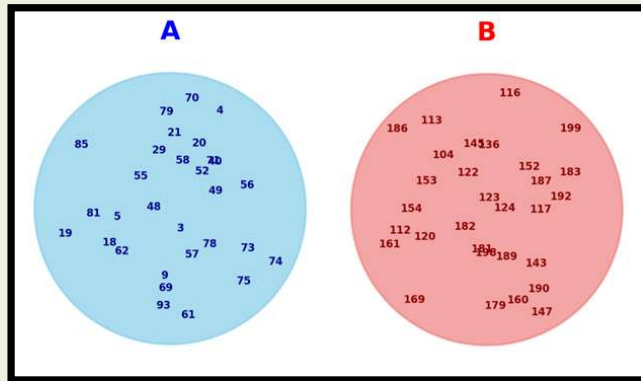
Nicel Değişken ile Bir Ortalama Değeri			
Parametrik		Nonparametrik	
Tek Örneklem T Testi		Tek örneklem K-S	
1 Nitel 1 Nicel Değişken			
Kategori sayısı 2		Kategori Sayısı 2'den çok	
Parametrik	Nonparametrik	Parametrik	Nonparametrik
Bağımsız Örneklem T Testi	Mann-Whitney U Testi	ANOVA	Kruskal Wallis H Testi
2 Nitel Değişken			
Bağımsız Ölçümler		Tekrarlı Ölçümler	
Ki-Kare Testi		McNemar Testi	
2 Nicel Değişken			
Normal dağılıma uygunluk (+)		Normal dağılıma uygunluk (-)	
Pearson Korelasyonu		Spearman Korelasyonu	
Tekrarlı Nicel Ölçümler			
Ölçüm (Değişken) Sayısı 2		Ölçüm (Değişken) Sayısı 2'den Çok	
Parametrik	Nonparametrik	Parametrik	Nonparametrik
Eşleştirilmiş Örneklem T Testi	Wilcoxon Testi	Tekrarlı Ölçümler İçin ANOVA	Friedman Testi

Parametrik Test Varsayımları:

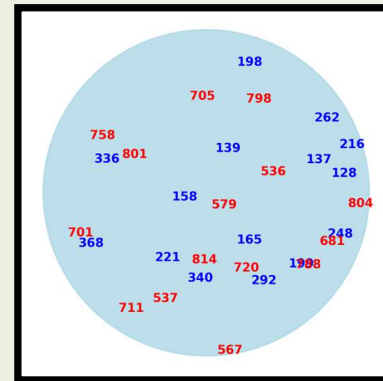
Örneklemin evrenden tesadüfi yollarla seçilmiş olması,
Gözlem sayısınının 30'dan büyük olması ($n > 30$),
Normal dağılıma uygunluk koşulu,
Varyansların homojenliği

Bağımlı ölçüm: Aynı örneklemden,
aynı ölçüm aracı kullanılarak
farklı zamanlarda elde edilmiş
veriler.

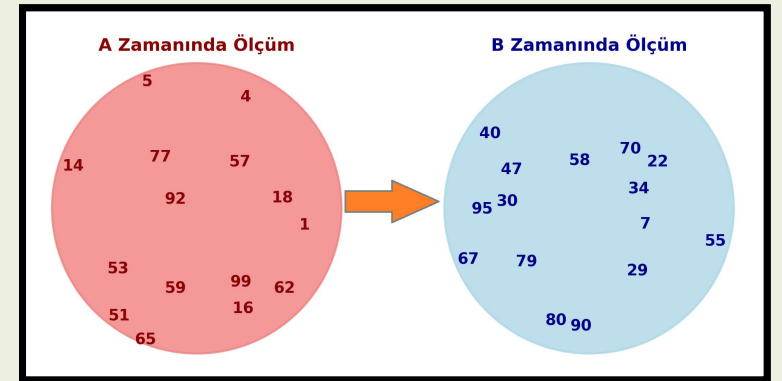
İLİŞKİ mi DEĞİŞİM mi FARKLILIK mı



A ve B kümeleri arasında **farklılık** yoktur/vardır



Kırmızı ve mavi ölçümler arasında **ilişki** yoktur/vardır



A zamanındaki ölçümlerle B zamanındaki ölçümler arasında **değişim** yoktur/vardır

İSTATİSTİK NE YAPAR?

- Doğru tanımlayıcı istatistik ve doğru yöntemden elde edilen p değerini sunar.
- p değeri 0,05'ten küçük değil ise "ilişki/değişim/farklılık yoktur" denir.
- p değeri 0,05'ten küçük ise tanımlayıcı istatistik yorumlanır.

Doğru Tanımlayıcı İstatistik nedir?

- Nitel-Nitel karşılaştırmalarında frekans ve yüzde
- Nitel-Nicel karşılaştırmalarında
 - *Parametrik yöntem - aritmetik ortalama \pm standart sapma*
 - *Nonparametrik yöntem – medyan (minimum-maksimum)*
- Nicel-Nicel karşılaştırmalarında korelasyon katsayısı

Dođru Tanımlayıcı İstatistik nedir?

Yaş <i>med (min-maks)</i>	25 (15-40)	Normal dağılıma uygunluk göstermediđi için med (min-maks)
Ağırlık <i>ort±SS</i>	67,74±18,81	Normal dağılıma uygunluk gösterdiđi için ort±SS

Nitel deđişken olduđu için frekans ve yüzde

Eđitim	f (%)
Ortaokul	156 (35)
Lise	8 (1,8)
Üniversite	145 (32,5)
Yüksek lisans	97 (21,7)
Doktora	40 (9)

BAĞIMSIZ ÖRNEKLEM T TESTİ

Anksiyete puanı			
Ek hastalık	$\bar{x} \pm SS$	t	p
Yok	11,47±4,00	-0,559	0,576
Var	11,76±4,15		

Ek hastalık durumu kategorileri arasında anksiyete puanları bakımından anlamlı farklılık olmadığı (p=0,576) tespit edilmiştir.
(Ek hastalığı olan ve olmayanların anksiyete puanları arasında anlamlı farklılık olmadığı (p=0,576) tespit edilmiştir.)

Cinsiyet	BKİ	t	p
	$\bar{x} \pm SS$		
Erkek	26,69±3,34	3,414	0,001*
Kadın	25,47±4,91		

Erkeklerin kadınlardan daha yüksek BKİ değerlerine sahip olduğu (p=0,001) tespit edilmiştir.

MANN-WHITNEY U TESTİ

Çalışma durumu	Maternal bağlanma		
	<i>Med (min – maks)</i>	Z	p
Çalışıyor	101 (86-103)	-0,667	0,505
Çalışmıyor	100 (87-104)		

Çalışma durumu kategorileri arasında maternal bağlanma puanları bakımından anlamlı farklılık olmadığı ($p=0,505$) tespit edilmiştir.

(çalışan ve çalışmayan annelerin maternal bağlanma puanları arasında anlamlı farklılık olmadığı ($p=0,505$) tespit edilmiştir.)

Bebek yoğun bakıma alındı mı	Maternal bağlanma		
	<i>Med (min – maks)</i>	Z	p
Evet	102 (96-105)	-2,721	0,007*
Hayır	97 (88-105)		

Bebeği yoğun bakıma alınan anneleri bebeği yoğun bakıma alınmayan kadınlardan daha yüksek maternal bağlanma puanlarına sahip olduğu ($p=0,007$) tespit edilmiştir.

TEK YÖNLÜ ANOVA

Boy (cm)			
	$\bar{x} \pm SS$	F	p
Egzersiz yapmıyorum (yapmıyordum)	169,33±7,27	2,117	0,097
1-30 dakika arası	170,95±7,68		
31-60 dakika arası	171,11±7,98		
61 dakika ve üzeri	169,39±9,27		

Günlük egzersiz süresi kategorileri arasında boy (cm) değerleri bakımından anlamlı farklılık olmadığı (p=097) tespit edilmiştir.

TEK YÖNLÜ ANOVA

	BKİ $\bar{x} \pm SS$	F	p
Egzersiz yapmıyorum (yapmıyordum)	26,87±4,11	3,602	0,013
1-30 dakika arası	26,56±3,53		
31-60 dakika arası	25,71±3,96		
61 dakika ve üzeri	25,54±4,83		

Günlük egzersiz süresi kategorileri arasında BKİ (cm) değerleri bakımından anlamlı farklılık olduğu (p=0,013) ve günlük 61 dakika ve üzeri egzersiz yapan bireylerin egzersiz yapmayan bireylerden daha düşük BKİ değerlerine sahip oldukları tespit edilmiştir.

Multiple Comparison **p değerleri**

Dependent Variable: VKİ
Tukey HSD

(I) Günlük_egzersiz_süre	(J) Günlük_egzersiz_süre	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Egzersiz yapmıyorum (yapmıyordum)	1-30 dakika arası	,30452	,45319	,908	-,8628	1,4719
	31-60 dakika arası	1,15395	,51614	,115	-,1755	2,4834
	61 dakika ve üzeri	1,32947*	,51085	,047	,0136	2,6453
1-30 dakika arası	Egzersiz yapmıyorum (yapmıyordum)	-,30452	,45319	,908	-1,4719	,8628
	31-60 dakika arası	,84944	,43231	,202	-,2641	1,9630
	61 dakika ve üzeri	1,02495	,42599	,077	-,0723	2,1222
31-60 dakika arası	Egzersiz yapmıyorum (yapmıyordum)	-1,15395	,51614	,115	-2,4834	,1755
	1-30 dakika arası	-,84944	,43231	,202	-1,9630	,2641
	61 dakika ve üzeri	,17552	,49243	,984	-1,0929	1,4439
61 dakika ve üzeri	Egzersiz yapmıyorum (yapmıyordum)	-1,32947*	,51085	,047	-2,6453	-,0136
	1-30 dakika arası	-1,02495	,42599	,077	-2,1222	,0723
	31-60 dakika arası	-,17552	,49243	,984	-1,4439	1,0929

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

kümeler

VKİ

Tukey HSD^{a,b}

Günlük_egzersiz_süre	N	Subset for alpha = 0.05	
		1	2
61 dakika ve üzeri	137	25,5364	
31-60 dakika arası	13	25,7120	25,7120
1-30 dakika arası	25	26,5614	26,5614
Egzersiz yapmıyorum (yapmıyordum)	114		26,8659
Sig.		,134	,071

Means for groups in homogeneous subsets are displayed.

- Uses Harmonic Mean Sample Size = 145,029.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Ortalama değerleri

KRUSKAL-WALLIS H TESTİ

Yaş	Günlük alınan protein (gr)		
	<i>Med (min – maks)</i>	H	p
18-22 Yaş	100 (89-103)	0,176	0,916
23-27 Yaş	100 (87-104)		
28 Yaş ve üzeri	99 (88-104)		

Yaş kategorileri arasında günlük alınan protein miktarları (gr) bakımından anlamlı farklılık olmadığı ($p=0,916$) tespit edilmiştir.


KRUSKAL-WALLIS H TESTİ



Yaş			
	Med (min – maks)	H	p
Büyükşehir	56 (18-89)	46,125	<0,001*
İl merkezi	43 (14-91)		
İlçe merkezi	34 (17-88)		
Belde-Köy-Kasaba	42,5 (19-88)		

Yerleşim yeri kategorileri arasında yaş değerleri bakımından anlamlı farklılık olduğu ($p < 0,001$) ve büyükşehirde yaşayan bireylerin ilçe merkezi ve il merkezinde yaşayan bireylerden daha yüksek yaşlara sahip olduğu tespit edilmiştir.

Each node shows the sample average rank of Yerleşim

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
İlçe merkezi-İl merkezi	48,046	30,294	1,586	,113	,677
İlçe merkezi-Belde-Köy-Kasaba	-62,133	40,020	-1,553	,121	,723
İlçe merkezi-Büyükşehir	139,188	24,677	5,640	,000	,000
İl merkezi-Belde-Köy-Kasaba	-14,087	38,100	-,370	,712	1,000
İl merkezi-Büyükşehir	91,143	21,424	4,254	,000	,000
Belde-Köy-Kasaba-Büyükşehir	77,056	33,805	2,279	,023	,136

 p değerleri

  Düzeltilmiş (Bonferroni düzeltilmeli) p değerleri

Bonferroni düzeltmesi: Testi güvenilirlik düzeyini kontrol etmek amacıyla yapılan bir düzeltmedir. Örnekte, yerleşim yeri kategorileri arasında yaş değerleri bakımından anlamlı farklılık olduğu tespit edilmiştir. Farklılığın hangi kategori ya da kategorilerden kaynaklandığını tespit etmek için kategoriler ikiyeşerli olarak karşılaştırılmıştır. Toplam ikiyeşerli karşılaştırma sayısı 6 olduğu için p değerleri (tabloda sig. sütununda yazan değerler) 6 ile çarpılarak düzeltilmiş p değerleri (adj. Sig. Sütununda yazan değerler) elde edilmiştir. Yorumlar, düzeltilmiş p değerleri üzerinden yapılmalıdır.

PEARSON & SPEARMAN KORELASYON KATSAYILARI

Boy		
Yaş	r	0,011
	p	0,786

Bireylerin yaşları ile boyları arasında anlamlı ilişki olmadığı ($p=0,786$) tespit edilmiştir.

BKİ		
Yaş	r	0,405
	p	<0,001

Bireylerin yaşları ile boyları arasında pozitif yönlü orta düzeyde ilişki olduğu ($p<0,001$; $r=0,405$) tespit edilmiştir.

Korelasyon katsayısı (r) nın yorumu

$p<0,05$ ise korelasyon katsayısının işaretine bakılır negatif ise negatif yönlü, pozitif ise pozitif yönlü;

- $r<0.2$ ise çok zayıf ilişki yada korelasyon yok
- 0.2-0.4 arasında ise zayıf ilişki
- 0.4-0.6 arasında ise orta şiddette ilişki
- 0.6-0.8 arasında ise yüksek düzeyde ilişki
- $0.8>$ ise çok yüksek düzeyde ilişki olduğu yorumu yapılır.

Kİ-KARE TESTİ

	Alb <3,5 n (%)	Alb ≥3,5 n (%)	Toplam n (%)	Ki-kare	p
Erkek	34 (30,4)	78 (69,6)	112 (100)	1,902	0,167
Kadın	21 (21,9)	75 (78,1)	96 (100)		
Toplam	34 (30,4)	78 (69,6)	112 (100)		

Erkek ve kadınlarda hipoalbuminemi görülme sıklıkları bakımından anlamlı farklılık olmadığı (p=0,167) tespit edilmiştir.

	KOAH değil n (%)	KOAH n (%)	Toplam n (%)	Ki-kare	p
Erkek	134 (33,9)	261 (66,1)	395 (100)	81,548	<0,001
Kadın	173 (70,6)	72 (29,4)	245 (100)		
Toplam	307 (48)	333 (52)	640 (100)		

Erkeklerde kadınlardan daha sık KOAH görüldüğü tespit edilmiştir (p<0,001).

MCNEMAR TESTİ

Başlangıç- 18.ay	Sigara(-) n (%)	Sigara (+) n (%)	Toplam n (%)	p
Sigara(-)	254 (99,6)	1 (0,4)	255 (100)	<0,001
Sigara (+)	107 (27,8)	278 (72,2)	385 (100)	
Toplam	361 (56,4)	279 (43,6)	640 (100)	

Bireylerin başlangıç ve 18. ay sigara kullanım durumları arasında anlamlı farklılık olduğu ($p<0,001$) ve sigara bırakanların oranının sigaraya başlayanların oranından daha yüksek olduğu tespit edilmiştir.

Başlangıç- 3.ay	Egzersiz (-) n (%)	Egzersiz (+) n (%)	Toplam n (%)	p
Egzersiz (-)	34 (30,4)	78 (69,6)	112 (100)	0,917
Egzersiz (+)	21 (21,9)	75 (78,1)	96 (100)	
Toplam	34 (30,4)	78 (69,6)	112 (100)	

Bireylerin başlangıç ve 3. ay egzersiz yapma durumları arasında anlamlı bir değişim olmadığı ($p=0,917$) tespit edilmiştir.

TEKRARLI ÖLÇÜMLER İÇİN ANOVA

	Type III Sum of Squares	Mean Square	F	p
Hemoglobin	35,75	17,875	0,877	0,416

Hemoglobin değerlerinde anlamlı bir değişim olmadığı saptanmıştır ($p=0,416$).

TEKRARLI ÖLÇÜMLER İÇİN ANOVA

	Type III Sum of Squares	Mean Square	F	p
BKI	41,56	25,84	11,006	<0,001

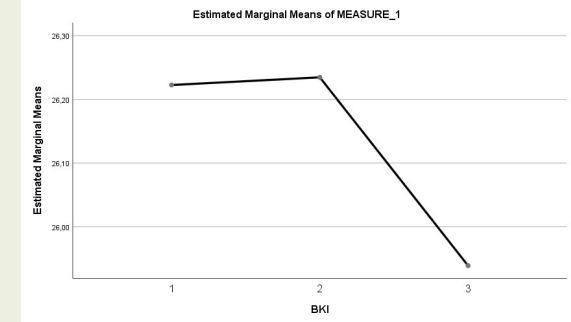
BKI değerlerindeki değişimler bakımından anlamlı farklılık olduğu ($p < 0,001$) tespit edilmiştir.

Farklılığın hangi ölçüm ya da ölçümlerden kaynaklandığının tespit edebilmek için ölçümler arasında ikişerli karşılaştırmalar yapılmıştır.

BKI	Type III Sum of Squares	Mean Square	F	p
Level 1 vs. Level 2	55,801	3,841	1,312	0,252
Level 2 vs. Level 3	19,154	4,154	9,054	0,001

Birinci ölçüm ile ikinci ölçüm arasında anlamlı bir değişim olmadığı ($p = 0,252$),

ikinci ölçümden üçüncü ölçüme kadar geçen sürede anlamlı bir azalış olduğu ($p = 0,001$) tespit edilmiştir.



FRIEDMAN TESTİ

	VKI (med-min-maks)	Hesap deęeri	p
VKI	25,21 (19,04-35,71)	4,929	0,085
Ay_3_VKI	25,32 (19,14-37,52)		
Ay_18_VKI	25,22 (20,19-27,21)		

- VKI ölçümleri arasında anlamlı deęişim olmadığı ($p=0,085$) tespit edilmiştir.

FRIEDMAN TESTİ

	VKI (med-min-maks)	Hesap değeri	p
VKI	26,29 (16,04-45,71)	18,144	<0,001*
Ay_3_VKI	26,05 (16,04-45,71)		
Ay_18_VKI	25,06 (16,04-45,71)		

- VKI ölçümleri arasında anlamlı değişim olduğu ($p < 0,001$) tespit edilmiştir.
anlamlı değişim olduğu için ölçümler Wilcoxon testi ile ikiyeşerli olarak karşılaştırılmıştır.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Ay_18_VKI-Ay_3_VKI	,155	,056	2,767	,006	,017
Ay_18_VKI-VKI	,506	,056	9,056	,000	,000
Ay_3_VKI-VKI	,352	,056	6,289	,000	,000

- *Karşılaştırmalar sonucunda; 3. ay VKİ değerlerinin ilk VKİ değerlerinden, 18. ay VKİ değerlerinin 3. ay VKİ değerlerinden daha düşük olduğu tespit edilmiştir.*

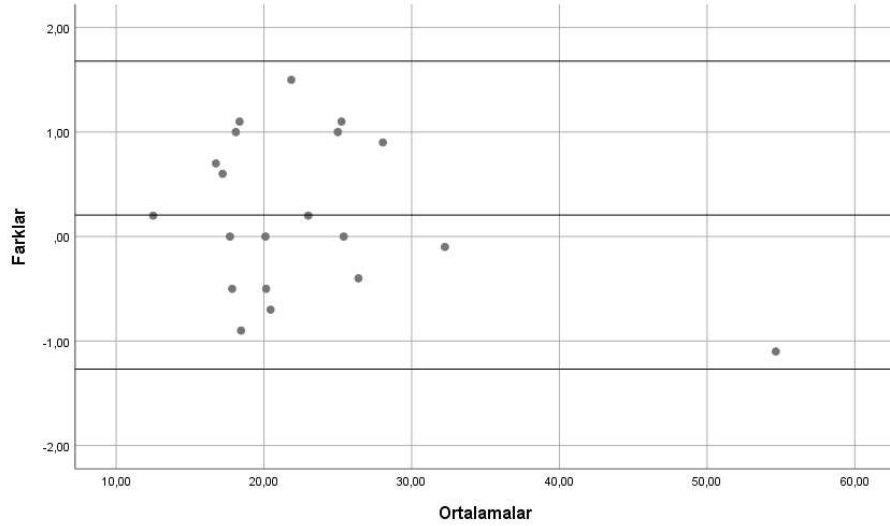
Uyum Analizleri

nicel ölçümler

Bland-Altman
Yöntemi

nitel ölçümler

Kappa
istatistiği



$p < 0,05$ ise uyum vardır
katsayı yorumlanır

KAPPA İSTATİSTİĞİ (TESTİ)

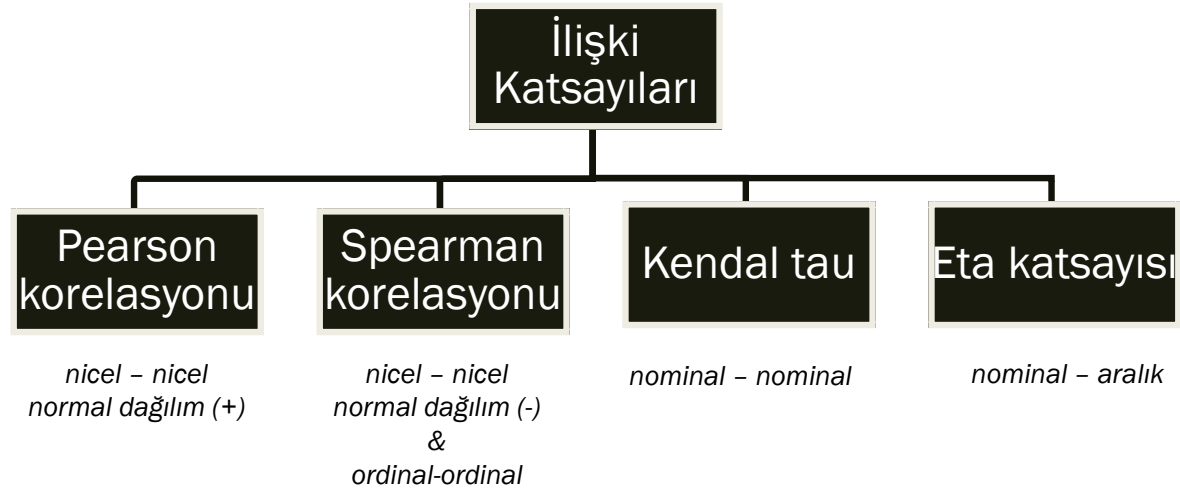
Hekim 1 / Hekim 2	Negatif	Pozitif	Kappa	p
Negatif	254 (99,6)	1 (0,4)	0,671	<0,001*
Pozitif	107 (27,8)	278 (72,2)		

Hakem 1 ve hakem 2 arasında önemli derecede uyum olduğu ($p < 0,001$; $K = 0,671$) tespit edilmiştir.

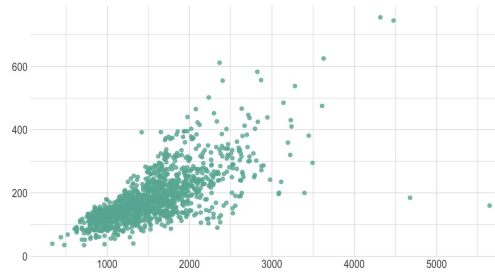
$p < 0,05$ ise;

- $K < 0$ uyum yok
- 0-0,20 önemsiz derecede uyum var
- 0,21-0,40 orta düzeyde uyum var
- 0,41-0,60 iyi uyum var
- 0,61-0,80 çok iyi derecede uyum var
- 0,81-1,00 mükemmel düzeyde uyum var

$p \geq 0,05$ ise uyum yok.



$p < 0,05$ ise katsayı yorumlanır

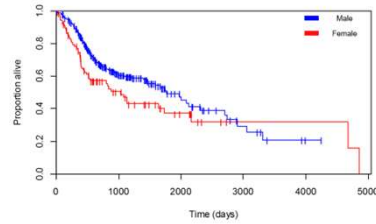


Sağkalım & tanı testleri

sağkalım analizi

Kaplan-Meier testi

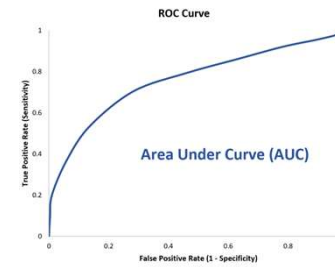
$p < 0,05$ ise kategoriler arasında sağkalım süreleri bakımından anlamlı farklılık vardır



tanı testleri

ROC analizi

$p < 0,05$ ise belirtilen biyomarker hastalık (ya da durumu) tanımlamada kullanılabilir



İnsanlık ve Makineler

- Makine: Herhangi bir enerji türünü başka bir enerjiye dönüştürmek, belli bir iş yapmak ya da bir etki oluşturmak için birleştirilmiş, genellikle çarklardan, dişlilerden, millerden ve daha birçok devinimli parçadan oluşan karmaşık düzenekler bütünü.
- Yapay zekâ: Bir bilgisayarın veya bilgisayar kontrollü robotun, genellikle akıllı varlıklarla ilişkili görevleri yerine getirme yeteneğidir.
- Makine Öğrenmesi: Sistemlerin verilerden öğrenebileceği, kalıpları belirleyebileceği ve minimum insan müdahalesi ile karar verebileceği fikrine dayanan yapay zekânın alt dalıdır.
- Derin Öğrenme: yapay sinir ağlarının (insan beyni gibi çalışacak şekilde modellenen algoritmalar) büyük miktarda veriden öğrendiği makine öğreniminin (ML) bir altkümesidir.

Tarihçe

- 1950-Turing Testi - Bir makine gerçekten düşünebilir mi?
- 1956: Dartmouth Konferansı - "yapay zeka" teriminin ilk kez kullanılması - John McCarthy
- 1957: Yapay sinir ağlarının kullanımı - Frank Rosenblatt
- 1958-1959: Makineler Düşünebilir mi ve Nasıl Düşünebilir? – Cahit Arf
- 1966: İlk Mobil Robot Shakey - Stanford Araştırma Enstitüsü – Otonom araba
- 1989: Evrimsel Ağların Doğuşu- Fransız bilim laboratuvarı – Posta kodu tanımlaması
- 1997: Deep Blue – IBM – Satranç – Garry Kasparov
- 2005: Stanley - SRI - ABD Savunma ve İleri Araştırma Projeleri Ajansı DARPA otonom araç yarışması şampiyonu

Tarihçe

- 2011: IBM Watson – Jeopardy bilgi yarışmasının en iyi iki oyuncusunu yener
- 2011: Siri – Apple – Sesli asistan
- 2014: Alexa – Amazon – Akıllı ev teknolojisinin önünün açılması
- 2016: AlphaGo – Google – Go oyunu şampiyonlarını yener
- 2016: Sophia - Hanson Robotics – görme, yüz ifadeleri yapma ve yanıt verebilme
- 2016: Google Home – Google – hatırlatma ve ses ile bilgi öğrenme

Üretken Yapay Zeka

- Üretken yapay zeka; yeni, orijinal içerik oluşturmaktan sorumlu bir yapay zeka türüdür.
- 2023: Chat GPT 4 – Open AI

İstatistiksel Öğrenme

Bağımlı değişken var

Bağımlı değişken yok

Bağımlı değişken nicel

Bağımlı değişken nitel

Sağkalım ve süreleri var

Küme sayısı bilinmiyor

Küme sayısı biliniyor

Doğrusal regresyon

Lojistik regresyon

Chaid analizi (karar ağacı)

Cox regresyon

Hiyerarşik Kümeleme

Hiyerarşik olmayan kümeleme

İki aşamalı kümeleme

Binary (ikili) lojistik Regresyon

Multinomial lojistik regresyon

Ordinal (sıralı) lojistik regresyon



1912-1954



1910-1997

DOĞRUSAL REGRESYON

Örnek: Bireylerin yaş ve kilo değerlerinin boy değerleri üzerine etkisinin incelenmesi.

		Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	149,379	1,659		90,063	,000	146,122	152,636
	Yaşınız	-,074	,016	-,173	-4,703	<,001	-,105	-,043
	Kilonuz	,324	,023	,524	14,220	<,001	,279	,369

a. Dependent Variable: Boyunuz

Yaş değerlerindeki artışla birlikte boy değerlerinde azalış meydana geldiği ($p < 0,001$), kilo değerlerindeki artışla boy değerlerinde de artış meydana geldiği ($p < 0,001$) tespit edilmiştir.

Yorumu nasıl yaptık?

Yaş ve kilo değerleri için p değerlerine (sig. Değerlerine) bakılır, anlamlı ise katsayıların (Unstandardized B değerlerinin) başındaki işaretleri incelenir. İşaret + ise bağımsız değişken artarken bağımlı değişken de artar, işaret – ise bağımsız değişken artarken bağımlı değişken azalır.

DOĞRUSAL REGRESYON

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
1	(Constant)	149,379	1,659	90,063	,000	146,122	152,636
	Yaşınız	-,074	,016	-,173	<,001	-,105	-,043
	Kilonuz	,324	,023	,524	<,001	,279	,369

a. Dependent Variable: Boyunuz

Tahmin denklemi nasıl oluşturulur?

Tahmin denklemi Unstandardized B değerleri yardımıyla oluşturulur.

Bağımlı değişkenimiz boy olduğu için;

$$\text{Boy}_{(\text{tahmin})} = 149,379 - 0,074 * \text{yaş} + 0,324 * \text{Kilo}$$

İKİLİ LOJİSTİK REGRESYON

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Cinsiyetiniz(1)	-,238	,382	,388	1	,533	,788	,373	1,667
	Yaşınız	,169	,020	69,588	1	,000	1,184	1,138	1,232
	VKİ	-,010	,048	,041	1	,840	,990	,902	1,088
	Full_Odun_tezek_kömür_süre	,066	,019	11,362	1	,001	1,068	1,028	1,109
	Sigara_binary(1)	2,320	,414	31,421	1	,000	10,175	4,521	22,899
	Yakınlarda_KOAH(1)	1,399	,394	12,627	1	,000	4,052	1,873	8,766
	Constant	-9,386	1,504	38,968	1	,000	,000		

Örnek'te KOAH olma durumunu (0: KOAH değil, 1: KOAH) etkileyen faktörlerin belirlenmesi amacıyla ikili lojistik regresyon gerçekleştirilmiştir. Bağımsız değişkenler aşağıdaki gibidir.

Cinsiyet (1:Erkek, 2:Kadın)

Yaş (Nicel)

VKİ (Nicel)

Odun tezek kömür dumanına maruz kalma süres (nicel-yıl)

Sigara_binary (0:Kullanmıyor, 1:Kullanıyor)

Yakınlarda KOAH (0:Yok, 1:Var)

İKİLİ LOJİSTİK REGRESYON

Variables in the Equation

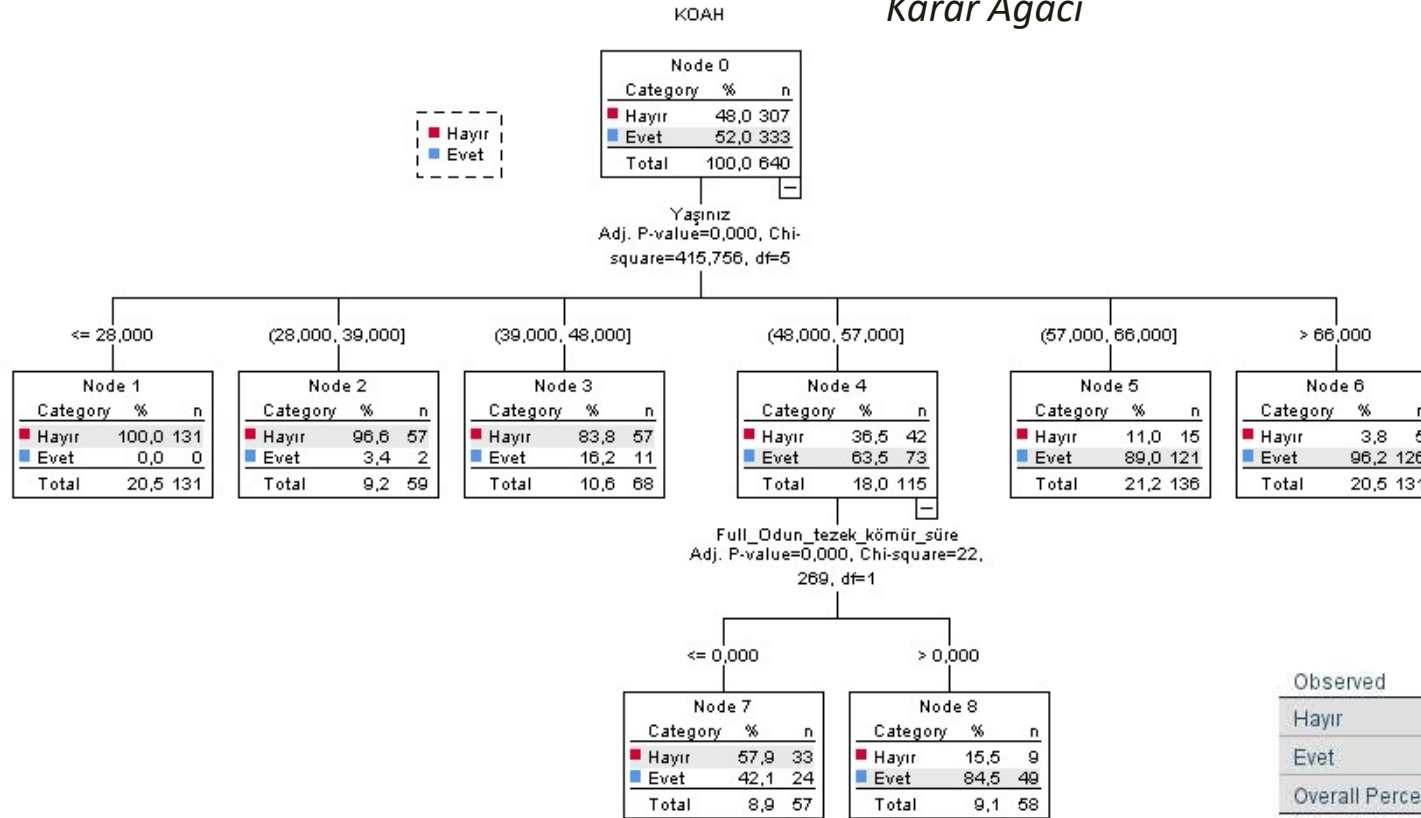
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Cinsiyetiniz(1)	-,238	,382	,388	1	,533	,788	,373	1,667
	Yaşınız	,169	,020	69,588	1	,000	1,184	1,138	1,232
	VKI	-,010	,048	,041	1	,840	,990	,902	1,088
	Full_Odun_tezek_kömür_süre	,066	,019	11,362	1	,001	1,068	1,028	1,109
	Sigara_binary(1)	2,320	,414	31,421	1	,000	10,175	4,521	22,899
	Yakınlarda_KOAH(1)	1,399	,394	12,627	1	,000	4,052	1,873	8,766
	Constant	-9,386	1,504	38,968	1	,000	,000		

Yorum:

Yaş değerlerindeki 1 birimlik artışın KOAH olma riskini 1,184 kat artırdığı, odun-tezek-kömür dumanına maruz kalma süresindeki 1 birimlik artışın KOAH olma riskini 1,068 kat artırdığı, sigara kullananların kullanmayanlara göre 10,175 kat daha yüksek riske sahip olduğu, yakınlarında KOAH olanların 4,052 kat daha yüksek riske sahip olduğu tespit edilmiştir.

CHAID ANALİZİ

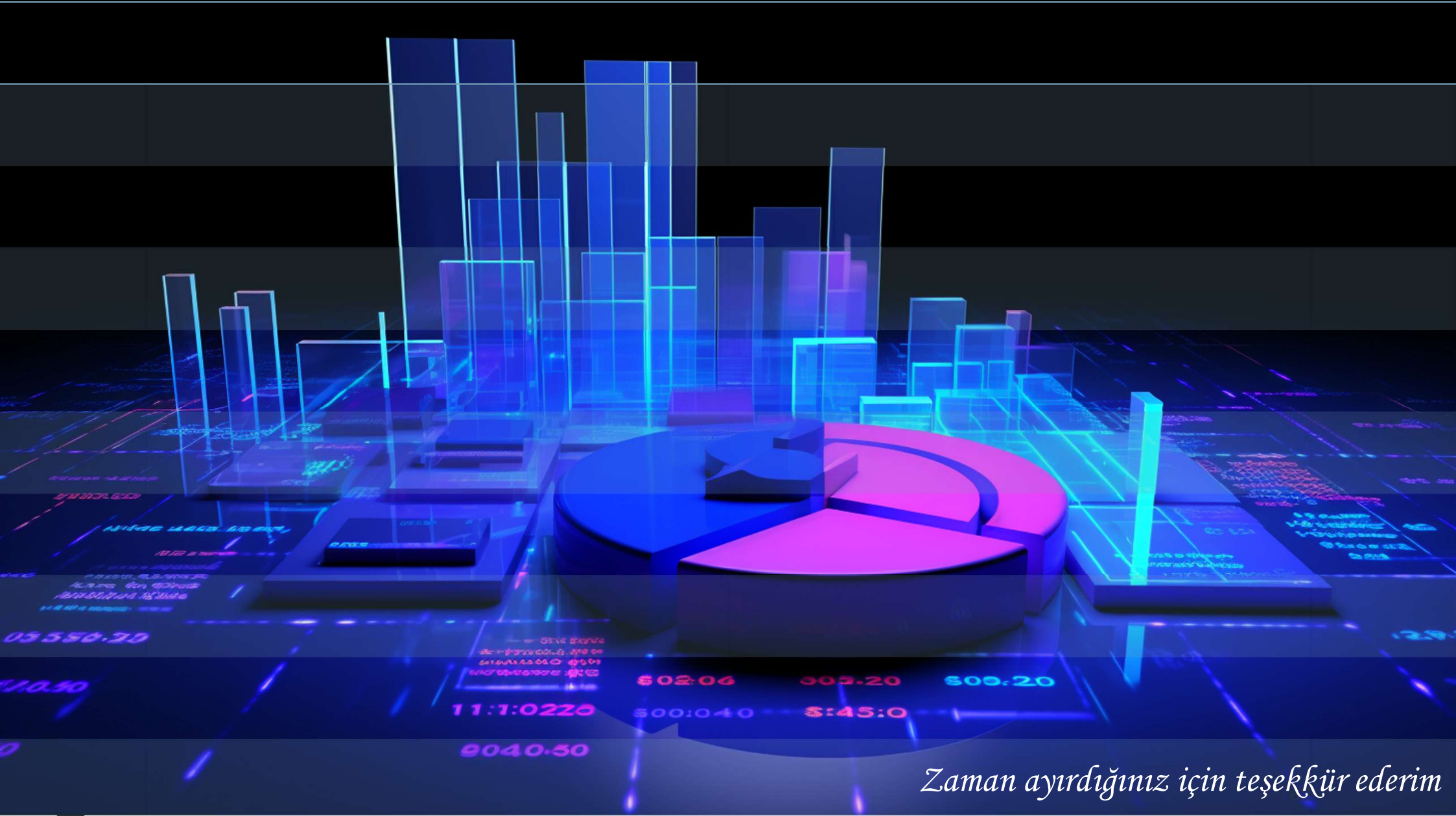
Karar Ağacı



Classification

Observed	Predicted		Percent Correct
	Hayır	Evet	
Hayır	278	29	90,6%
Evet	37	296	88,9%
Overall Percentage	49,2%	50,8%	89,7%

Growing Method: CHAID
Dependent Variable: KOAH



Zaman ayırdığınız için teşekkür ederim