

# **KORELASYON VE REGRESYON ANALİZİ**

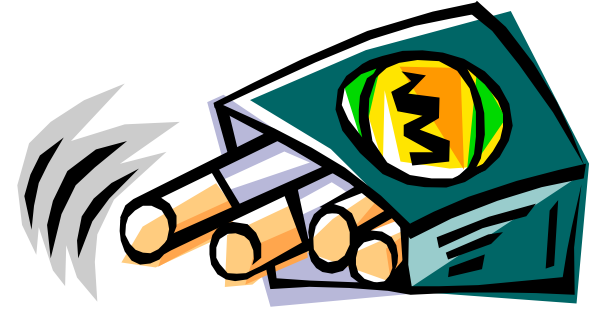
**Ankara Üniversitesi Tıp Fakültesi  
Biyoistatistik Anabilim Dalı**

İki ya da daha çok deęişken arasında ilişki olup olmadığını, ilişki varsa yönünü ve gücünü inceleyen “korelasyon analizi” ile deęişkenlerden birisi belirli bir birim deęiştğinde dięerinin nasıl bir deęişim gösterdiğini inceleyen “regresyon analizi” saęlık bilimlerinde çok kullanılan istatistiksel yöntemlerdir.

## *Korelasyon ve regresyon analizinin sađlık bilimlerinde kullanımına iliřkin bazı rnekler:*

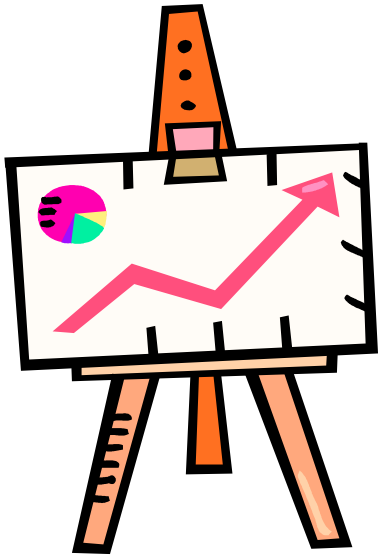
Hastalıđın bařlangıcıyla doktora bařvurma arasında geen sre ile iyileřme sresi arasındaki, yine ilacın dozu ile iyileřme sresi arasındaki iliřkilerden yararlanılarak tedaviye bařlama, dozunu ayarlama ve tedaviyi bitirmeye karar verilir.



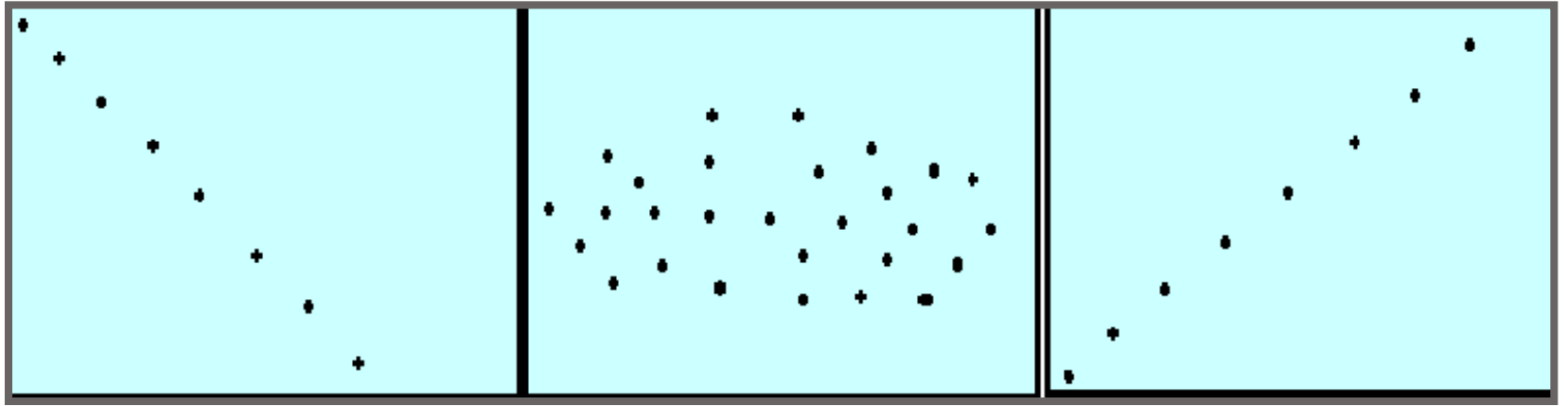


- \* Sigara içme sıklığı ile koroner kalp hastalığına yakalanma arasında nedensel bir ilişki saptanırsa sigara içme önlenerek koroner kalp hastalığına yakalanma ortadan kaldırılabilir.

- Değişkenler arasındaki ilişkilerden yararlanılarak geliştirilecek matematiksel modeller yardımıyla tahminler yapılabilir. Örneğin;



- Yaş ile boy arasında bir model geliştirilebilirse bir çocuğun belirli yaşlarda boyunun kaç cm. olacağı,
- Doğum sayısı ile hemoglobin düzeyi arasında bir model geliştirilebilirse belirli sayıda doğum yapacak bir kadının hemoglobin değerinin ne olacağı,
- Belirli bir hastalıkta kullanılan bir ilacın dozu ile iyileşme süresi arasında bir model geliştirilebilirse verilen doza göre hastanın ne kadar sürede iyileşebileceği tahmin edilebilir.



$r = -1$   
Mükemmel  
negatif ilişki

$r = 0$   
İlişki yok

$r = +1$   
Mükemmel  
pozitif ilişki

Saçılım grafikleri iki değişken arasındaki ilişki hakkında genel bir bilgi edinmemizi sağlar. Ancak, ilişkinin miktarı konusunda yorum yapabilmek için korelasyon katsayısı hesaplamamız gerekmektedir.

# KORELASYON KATSAYISI

Korelasyon katsayısı ( $r$ ), iki deęişken arasındaki ilişkinin ölçüsüdür. Korelasyon katsayısı  $-1$  ve  $+1$  arasında deęişim gösterir.

$r=0$ ; iki deęişken arasında ilişki olmadığı,  
 $r=1$ ; mükemmel pozitif doğrusal ilişki olduğu,  
 $r=-1$ ; mükemmel negatif doğrusal ilişki olduğu

anlamına gelir.

Korelasyonun katsayısının gücü ile ilgili olarak aşağıdaki tanımlamalar yapılmıştır:

0.00 - 0.25 Çok zayıf ilişki

0.26 - 0.49 Zayıf ilişki

0.50 - 0.69 Orta ilişki

0.70 - 0.89 Yüksek ilişki

0.90 - 1.00 Çok yüksek ilişki



**Korelasyon katsayıları, aralarında  
ilişki araştırılan değişkenlerin  
türlerine göre farklılık gösterir.**

## *Sınıflanabilir nitel değişkenler arasındaki ilişkinin belirlenmesinde kullanılan korelasyon katsayıları*

- Phi Katsayısı

(Değişkenlerin her ikisi de 2 kategorili ise)

- Cramer V Katsayısı
- Olağanlık Katsayısı
- Lambda Katsayısı

## *Sıralanabilir nitel değişkenler arasındaki ilişkinin belirlenmesinde kullanılan korelasyon katsayıları*

- Spearman Korelasyon Katsayısı
- Gamma Katsayısı
- Kendall'in tau-b Katsayısı
- Kendall'in tau-c Katsayısı
- Somer'in d Katsayısı

## *Kesikli/Sürekli Nicel değişkenler arasındaki ilişkinin belirlenmesinde kullanılan korelasyon katsayıları*

- Pearson Korelasyon Katsayısı (Değişkenlerin her ikisi de normal dağılım gösteriyorsa)
- Spearman Korelasyon Katsayısı (Değişkenlerden en az birisi normal dağılım göstermiyorsa)

*Sınıflanabilir Nitel bir deęişken ve Kesikli/Sürekli bir Nicel deęişken arasındaki ilişkinin belirlenmesinde kullanılan korelasyon katsayıları*

- Çift Serili Korelasyon Katsayısı
- Nokta Çift Serili Korelasyon Katsayısı

*Sıralanabilir Nitel bir deęişken ve Kesikli/Sürekli bir Nicel deęişken arasındaki ilişkinin belirlenmesinde kullanılan korelasyon katsayısı*

- Çoklu Serili Korelasyon Katsayısı

# DOĞRUSAL REGRESYON

Regresyon analizi, bilinen bulgulardan, bilinmeyen gelecekteki olaylarla ilgili tahminler yapılmasına izin verir. Regresyon, bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi ve doğrusal eğri kavramını kullanarak, bir tahmin eşitliği geliştirir. Değişkenler arasındaki ilişki belirlendikten sonra, bağımsız değişken(ler)in değeri bilindiğinde bağımlı değişkenin değeri tahmin edilebilir.

## Bağımlı Değişken (y)

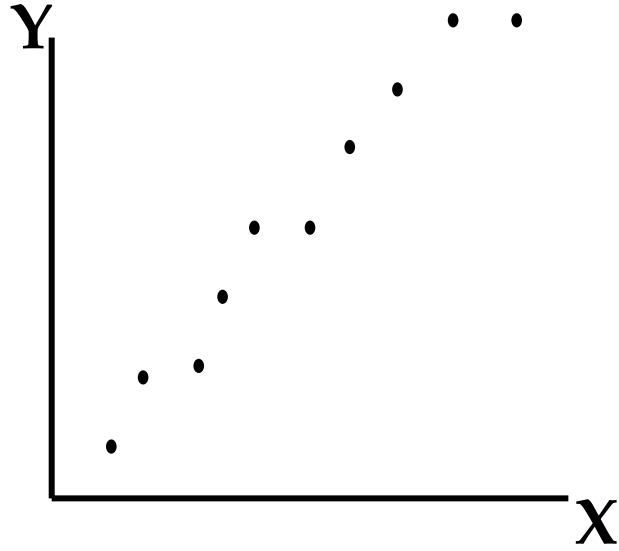
- Bağımlı değişken, regresyon modelinde açıklanan ya da tahmin edilen değişkendir. Bu değişkenin bağımsız değişken ile ilişkili olduğu varsayılır.

## Bağımsız Değişken (x)

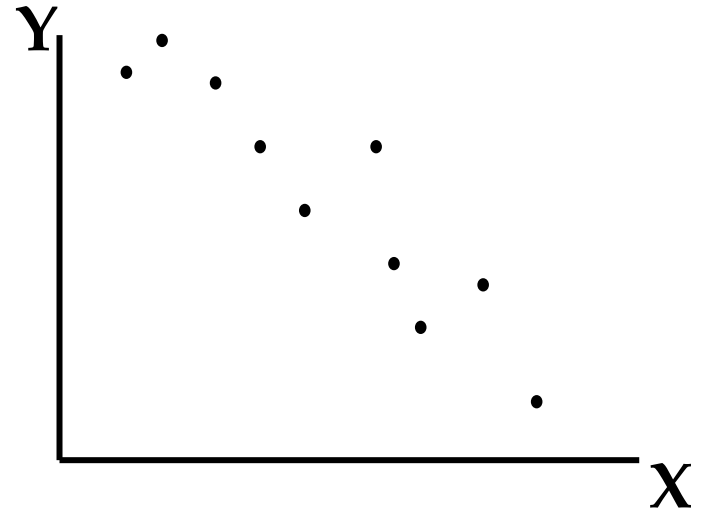
- Bağımsız değişken, regresyon modelinde açıklayıcı değişken olup; bağımlı değişkenin değerini tahmin etmek için kullanılır.

- Değişkenler arasında doğrusal ilişki olabileceği gibi, doğrusal olmayan bir ilişki de olabilir. Bu nedenle, saçılım grafiği yapılmadan (ilişki yok/doğrusal ilişki var/doğrusal olmayan ilişki var) regresyon analizine karar verilmemesi gerekir.
- Bu bilgiler doğrultusunda, tek/çok değişkenli doğrusal regresyon analizlerinin yanı sıra, tek/çok değişkenli doğrusal olmayan regresyon analizleri de mevcuttur.

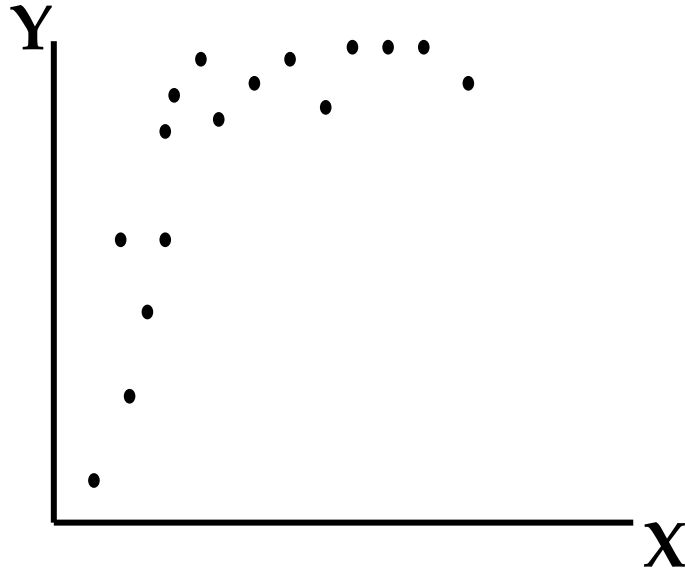




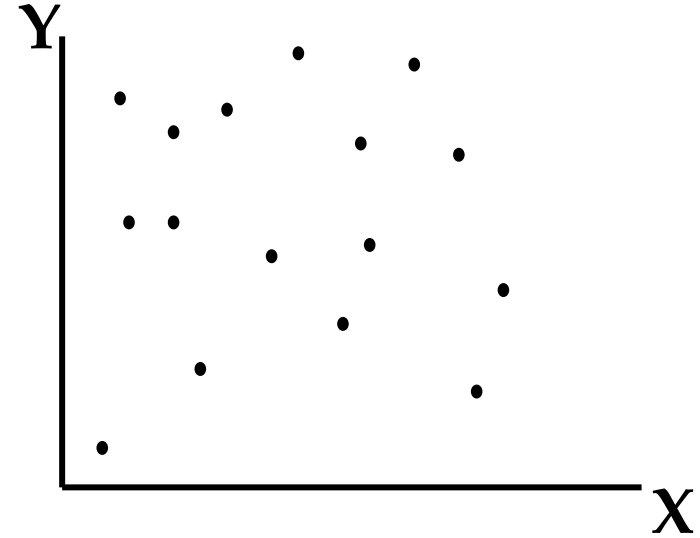
(a) (+) yönlü Doğrusal İlişki



(b) (-) yönlü Doğrusal İlişki



(c) Doğrusal Olmayan İlişki



(d) İlişki Yok

Bağımlı değişken ile bağımsız değişken arasındaki doğrusal ilişkiyi açıklayan tek değişkenli regresyon modeli aşağıdaki gibidir:

$$y = a + bx$$

**y** = Bağımlı değişkenin değeri

**a** = Regresyon doğrusunun kesim değeri (Sabit değer)

**b** = Regresyon doğrusunun eğimi

**x** = Bağımsız değişkenin değeri

## Örnek:

Kardiyoloji kliniğine başvuran 20 yaş üzerindeki erkek hastalar üzerinde yapılan bir araştırmada, yaş ve kolesterol değişkeni arasındaki ilişkiden yola çıkılarak kurulan regresyon modeli aşağıdaki gibi elde edilmiştir:

$$Y = 151.44 + 1.428X$$

$$\text{Kolesterol} = 151.44 + 1.428x\text{Yaş}$$

Bu modele göre, yaştaki bir birimlik artışın, kolesterol değerinde 1.428 birimlik bir artışa neden olacağı, 20 yaşındaki bir erkeğin ( $X=20$ ) kolesterol değerinin ise 180 olacağı söylenebilir.

Kurulan bu modele göre, 50 yaşında bir erkeğin kolesterol değerinin ne kadar olacağını tahmin edebiliriz

X=50 için :

$$Y = 151.44 + 1.428X$$
$$= 151.44 + 1.428 * 50 = 222.84$$

50 yaşında bir erkeğin kolesterol değerinin 222.84 olacağı söylenebilir.

## *BELİRTME KATSAYISI ( $R^2$ )*

Belirtme katsayısı, doğrusal modelin uyum iyiliğinin en iyi ölçüsüdür. Söz konusu katsayı, bağımlı değişkendeki değişimin ne kadarının bağımsız değişken(ler) tarafından açıklandığını ifade eder. Bu durum, regresyon modelinin açıklayıcılık gücünün iyi bir göstergesidir.

Yukarıdaki örneğe göre, belirtme katsayısı %74.3 olarak bulunmuştur. Bunun anlamı, yaş değişkeni kolesterol değişkenindeki değişimin %74.3'ünü açıklamaktadır. Geriye kalan %25.7'lik değişim, modele katılacak başka değişkenler tarafından açıklanabilir.



Dođrusal ve dođrusal olmayan regresyon modellerinin dıřında deđiřik regresyon analizleri de bulunmaktadırdır.

# Kanonik Korelasyon

Birden fazla sayıda bağımlı değişken ile bir ya da daha fazla bağımsız değişken arasındaki ilişkinin incelenmesi için kullanılan bir yöntemdir.

*Örneğin, HDL, LDL, Trigliserid değerlerini etkileyen değişkenlerin araştırılmasında kanonik korelasyon yönteminden yararlanılır.*



# Lojistik Regresyon

Bağımlı değişken iki sonuçlu bir değişken (Var/Yok, Hasta/Sağlıklı vb.) olduğunda, bu sonucun birinin (Var veya Yok) ortaya çıkma olasılığı üzerinde etkili olduğu düşünülen faktör(ler)in araştırılması yöntemidir.

*Örneğin; meme kanseri riski (Kanser / Kanser değil) üzerinde etkili olan risk faktörlerinin araştırılmasında lojistik regresyon yönteminden yararlanır.*

# Cox Regresyon

Zaman içinde izlenmiş ve belli tip bir sonuca erişmiş/erişmemiş olguların (iki sonuçlu değişken) değerlendirildiği ve bu sonucu oluşturmada çeşitli değişkenlerin etkilerinin araştırıldığı çözümlene yöntemidir.

*Örneğin, operasyon sonrası sağ kalım üzerinde etkili olan değişkenlerin araştırılmasında Cox regresyon yönteminden yararlanılır.*