

# İstatistik I

14. Hafta

4 Kasım 2021

14. Regrasyon ve Korelasyon

[www.umitsarp.com](http://www.umitsarp.com)

**Ümit SARP, [umit.sarp@ikcu.edu.tr](mailto:umit.sarp@ikcu.edu.tr)**

# Giriş

- Regrasyon
- Korelasyon



# Giriş

İki değişken arasındaki ilişkiyi tanımlamaya ve ölçmeye **korelasyon analizi** adı verilir.

Bir veya birden çok değişkenin başka bir değişken üzerindeki ilişkisini açıklamaya **regresyon analizi** adı verilir.

Yukarıdaki tanımlar her ne kadar birbirine benzer gözükse de korelasyon analizi ve regresyon analizi bir bozuk paranın iki farklı yüzü gibidirler.



# Giriş

Regresyon terimi 19. yüzyılda İngiliz istatistikçisi Francis Galton tarafından bir biyolojik inceleme için ortaya atılmıştır.

Galton babaların boyları ile oğullarının boyları arasındaki ilişkiyi araştırmıştır.

Araştırmaları sonucunda ortalamaya doğru bir eğilimin varlığını fark etmiştir.



# Giriş

Çok kısa boylu babaların oğullarının boylarının ortalama değerler etrafında (babalarından daha uzun gibi) toplandığını gözlerken bu durumun tersinin de doğru olduğunu (uzun boylu babaların oğulları da ortalama boyda bulunmaktadır) fark etmiştir.

Galton oğullarının boylarının ortalamaya doğru yönlendiğini (İngilizce karşılığı “regressed” olmak üzere) belirterek regresyon kelimesinin temelini de atmıştır.

Günümüzde regresyon kelimesi iki ya da daha fazla değişken arasındaki ilişkinin modellenmesi işlemlerinin tümünü içeren geniş bir anlama sahiptir.



# Giriş

Regresyon analizinde iki farklı değişken tanımlamasına ihtiyaç duyulmaktadır: Bunlar sırasıyla bağımlı ve bağımsız değişken kavramlarıdır.

Bağımlı değişken araştırmacının üzerinde çalıştığı değişken olup bu değişken üzerinde meydana gelen değişimlerin ya da bu değişkenin toplam değişkenliğinin açıklanılmasına çalışılmaktadır.

Bağımsız değişken ya da değişkenler ise ilgilenilen bağımlı değişkende meydana gelen değişim ya da toplam değişkenliğinin üzerinde etkisi olabileceği düşünülen değişken ya da değişkenlerdir.



# Giriş

Örneğin bağımlı değişken trafik akım hızı ise bağımsız değişkenler birim otomobil cinsinden trafik hacminin kapasiteye oranı, birim otomobil cinsinden trafik hacmi içindeki bisiklet sayısı, yol üzerindeki ticari yoğunluk oranı, güzergah üzerinde bulunan sinyalli ve sinyalsiz önemli kavşak sayısı olabilir.

Regresyon analizinde bir ya da daha fazla bağımsız değişken olabilir.

Bu ünite içerisinde tek bağımsız değişken olması durumu basit doğrusal regresyon analizi olarak ele alınacak, birden fazla bağımsız değişken olması durumu çoklu doğrusal regresyon analizi ele alınmayacaktır.



# Giriş

İlgilenilen iki değişken arasındaki ilişkinin derecesi için korelasyon analizi kullanılır.

Korelasyon analizinin regresyon analizinden farklılık gösterdiği nokta, korelasyon analizinin değişkenler arasındaki ilişkinin yalnızca derecesini göstermesidir.

İki değişken arasında yüksek korelasyon olması bu iki değişkenden birinin diğerinin nedeni olabileceğini göstermez. Korelasyon analizi iki değişken arasındaki nedensellik için kullanılmaz.

Nedensellik araştırması için farklı istatistik tekniklerinin kullanılması gerekir.





# Korelasyon Analizi

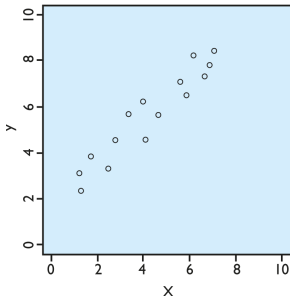
## Korelasyon Analizi:

En genel anlamı ile iki deęişken arasındaki ilişkinin **tanımlanması** ve ilişkinin **derecesinin belirlenmesi**dir. Deęişkenler arasında var olabilecek ilişkinin derecesinin tespit edilebilmesi amacı ile çeşitli teknikler kullanılabilir. En basit şekli ile iki deęişken arasındaki ilişkiyi gözlemlemek için bu deęişkenlerin dağılım **grafikleri çizilebilir**.

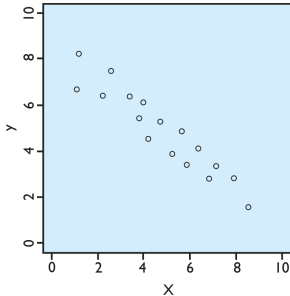
Aşağıdaki şekiller'de iki deęişken arasında gözlemlenebilecek dört farklı durum örneklenmiştir.



# Korelasyon Analizi



(a)

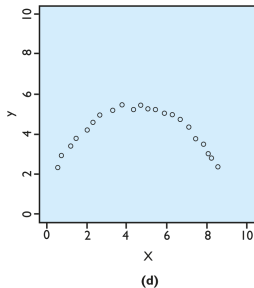
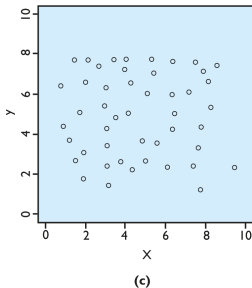


(b)

Şekil (a) ve (b) grafiklerinde yer alan iki değişkenin arasındaki ilişkinin derecesi birbirine eşittir, fakat iki değişken arasındaki ilişkinin yönü farklıdır. Dikkat edilirse Şekil (a)'da x değişkeninin değeri artarken y değişkeninin değeri de artmaktadır. Ancak Şekil (b)'de x değişkeninin değeri artarken y değişkeninin değeri azalmaktadır.



# Korelasyon Analizi



Şekil (c) ve (d) grafiklerinde yer alan iki değişkenin dağılım grafikleri incelendiğinde Şekil (c) grafiğinde yer alan değişkenler arasındaki ilişkinin rassal olduğu gözlemlenmektedir. Bu nedenle korelasyon analizi yapıldığında aralarında bir ilişki çıkmayacaktır. Şekil (d) dağılım grafiğinde ise iki değişken arasındaki ilişkinin eğrisel olduğu görülmektedir



# Korelasyon Analizi

İki deęişken arasındaki ilişkinin yalnızca grafikler ile incelenmesi yeterli olmayacaktır.

İlişkinin derecesini gösteren istatistiklere ihtiyaç duyulmaktadır.

İlişkinin derecesi için hesaplanacak **Pearson korelasyon katsayısı** incelenmiştir.



# Korelasyon Analizi

## Pearson Korelasyon Katsayısı:

İki ya da daha fazla, oranlı ve eşit aralıklı ölçeğe uygun şekilde ölçümlenmiş değişkenler arasındaki ilişkinin derecesini belirlemek için Pearson korelasyon katsayısı kullanılır.

Evrendeki  $x$  ve  $y$  değişkenlerinin korelasyon değeri  $\rho$  ile sembolize edilir ve bu sembol “ $\rho$ ” şeklinde okunur. Örneklemdaki  $x$  ve  $y$  değişkenlerinin korelasyon değeri  $r$  simgesi ile gösterilir.

Yapılan çalışmalar genellikle örneklemeler üzerinden gerçekleştirildiğinden bu ünite içinde **Pearson korelasyon katsayısı sembolü olarak  $r$**  kullanılmıştır.



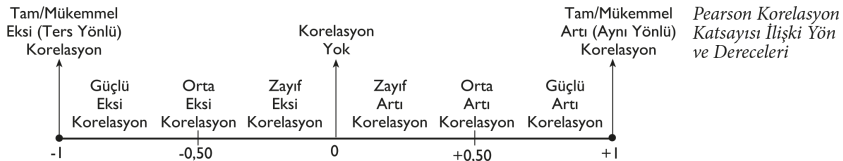
# Korelasyon Analizi

Pearson korelasyon katsayısı  $r$ ,  $-1$  ile  $+1$  arasında değişen değerler almaktadır. Pearson korelasyon katsayısı  $r$ 'nin  $-1$  ve  $+1$  değerlerine eşit sonuçlar, mükemmel/tam ilişkinin varlığını gösterir. Pearson korelasyon katsayısının eksi  $r$  değerleri değişkenler arasındaki ters yönlü ilişkiyi gösterirken (biri artarken diğ erinin azalması gibi), artı  $r$  değerleri değişkenler arasındaki aynı yönlü ilişkinin (biri artarken diğ eri de artmaktadır gibi) var olduğunu gösterir.

Eğ er iki değişken arasında hiç ilişki yok ise bir başka ifade ile değişkenler bağımsız ise Pearson korelasyon katsayısı  $0$  (sıfır) değerini alır. Şekil 8.1 (c) değişkenler arasında ilişkinin olmadığını bir başka ifade ile değişkenler arasındaki korelasyonun sıfır değerine eşit olduğunu göstermektedir.



# Korelasyon Analizi



Perason korelasyon katsayısı  $r$ ,  $-1$  ve  $+1$  değerlerine yaklaştıkça ilişkinin derecesinin arttığı ifade edilirken sıfır değerine yaklaştıkça ilişkinin derecesinin azaldığı/zayıfladığı ifade edilir.  $r$  değeri  $-0,50$  ya da  $0,50$  etrafında bir değer ise değişkenler arasında orta düzeyli bir ilişkinin varlığı ifade edilir. Şekil 8.2'de Pearson korelasyon katsayısı  $r$  için ilişkinin derecesi ve yönü özetlenmektedir.



# Korelasyon Analizi

Bu hesaplamaların yürütülmesi zaman alacağından Pearson korelasyon katsayısı ortalama ve standart sapmalara ihtiyaç duyulmaksızın,

$$r = \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}}$$

eşitliği yardımıyla da hesaplanabilir. Her iki eşitlikte aynı sonucu verecektir. Orijinal verinin var olması durumunda verilen ikinci eşitlik hesaplama kolaylığı sağlamaktadır.





# Korelasyon Analizi

## Belirlilik Katsayısı:

Pearson korelasyon katsayısı yardımıyla değişkenler arasındaki ilişkinin derecesi belirlenmiş ve bu bilgiyi kullanarak ilişkinin zayıf, orta ilişki gibi nitelmesini yapmıştık. Araştırmacılar, bağımlı değişkende meydana gelen değişim içerisinde bağımsız değişkenin payının ne olduğunu bilmek isteyebilir. Bu bilgiyi gösteren istatistiğe **belirlilik katsayısı** adı verilir.



# Korelasyon Analizi

Evrendeki  $x$  ve  $y$  değişkenlerinin belirlilik katsayısı  $\rho^2$  ile sembolize edilir.

Örneklemdaki  $x$  ve  $y$  değişkenlerinin belirlilik katsayısı  $r^2$  simgesi ile gösterilir.

Araştırmalar genellikle örneklemeler üzerinden yapıldığından bu ünite içinde belirlilik katsayısının simgesi olarak  $r^2$  kullanılacaktır.

Belirlilik katsayısı, Pearson korelasyon katsayısı  $r$ 'nin karesinin alınması ile hesaplanır ve  $r^2$  ile gösterilir.

0 ile 1 arasında değerler alır ve oran olarak ifade edilir.



# Korelasyon Analizi

## Korelasyon Katsayısı Anlamlılık Testi:

Arařtırmalarda çoęunlukla rneklemeler zerinden alıřılmaktadır. Bir arařtırmada rneklem sonularına gre belirlenen korelasyon deęerinin evren deęeri iin test edilmesi istenebilir. rneklem sonularına gre “Korelasyon yoktur.” ynnde karar verildiyse bu kararın evren iinde geerli olup olmadıęı korelasyon katsayısı anlamlılık testi yardımıyla yapılır. Korelasyon katsayısının anlamlılık testi, ařaęıdaki adımsal srete gerekleřtirilebilir.



# Korelasyon Analizi

## Adım 1: Hipotezlerin ifade edilmesi:

Korelasyon katsayısına ilişkin hipotezler aşağıdaki gibi ifade edilir:

$H_0 : \rho = 0$  ( $x$  ve  $y$  değişkenleri arasında korelasyon yoktur.)  $H_1 : \rho \neq 0$  ( $x$  ve  $y$  değişkenleri arasında korelasyon vardır.)



# Korelasyon Analizi

## Adım 2: Anlamlılık düzeyinin belirlenmesi:

Anlamlılık düzeyinin belirlenmesi, doğru olan sıfır hipotezinin, örneklemden elde edilen bilgilere dayanarak reddedilmesi olasılığını belirleyen  $\alpha$ 'nın seçilmesidir.  $\alpha$  anlamlılık düzeyi, araştırmacı tarafından, hipotezler ifade edilip veri derlenmeye başlanmadan önce seçilmelidir. Sosyal bilim araştırmalarında  $\alpha$  için genellikle %5 ve %1 değerleri seçilmektedir. Yapılan bu seçimle birlikte, doğru olan  $H_0$  hipotezinin reddedilme olasılığı belirlenmiş olur. Bu olasılık örnekleme dağılımıyla ilişkilendirilerek kullanılır. Bu durumda,  $\alpha$  anlamlılık düzeyi, doğru olan sıfır hipotezinin reddedilmesi olasılığına eşit olan örnekleme dağılımındaki oransal alanı göstermiş olur.



# Korelasyon Analizi

## Adım 3: İstatistiksel Test:

Testin gerçekleştirilmesinde  $n - 2$  serbestlik derecesi ile t dağılımı tablosundan elde edilen kritik değerler kullanılır. Daha sonra belirlenen anlam düzeyine göre t tablosu yardımıyla kritik değer tespit edilir. Örneklem korelasyon değeri yardımıyla,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

istatistiği hesaplanılarak tablodan elde edilen kritik değer ile karşılaştırılır.



# Korelasyon Analizi

## Adım 4: İstatistiksel kararın verilmesi:

Eğer hesaplanan  $t$  değeri,  $t$  tablosu yardımıyla belirlenen kritik değer (– kritik değer, + kritik değer) aralığında yer alıyor ise  $H_0$  ( $x$  ve  $y$  değişkenleri arasında korelasyon yoktur.) hipotezi kabul edilir. Bu aralık dışında yer alıyor ise  $H_0$  ( $x$  ve  $y$  değişkenleri arasında korelasyon yoktur.) hipotezi reddedilir,  $H_1$  ( $x$  ve  $y$  değişkenleri arasında korelasyon vardır.) hipotezi kabul edilir.



# Basit Doğrusal Regresyon Analizi

Önceki kısımlarda Pearson korelasyon katsayısı yardımıyla iki değişken arasındaki doğrusal ilişkinin derecesi tespit edilmiştir.

İki değişken arasındaki ilişki bir modelle açıklanmak istenebilir.

İstatistiksel anlamda iki değişken arasındaki ilişki denildiğinde, bu değişkenlerin değerlerinin karşılıklı değişimleri arasında bir neden sonuç ilişkisi veya bir bağımlılık ilişkisi anlaşılır.

Regresyon kelimesi, bir değişkenle bir başka değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır.





# Basit Doğrusal Regresyon Analizi

Bağımsız değişken  $x$ 'in değerleri değişirken, buna bağlı olarak bağımlı değişken  $y$ 'nin değerleri de değişiyorsa bu iki değişken arasında ilişki olduğu söylenebilir.

Değişkenler arasındaki bu ilişkinin matematiksel bir fonksiyonla ifade edilmesi, **regresyon analizinin konusunu** oluşturur.

İki değişken arasındaki ilişkinin gösteriminde kullanılan fonksiyona veya eşitliğe **regresyon denklemi** adı verilir.



# Basit Doğrusal Regresyon Analizi

Değişkenler arasındaki ilişki için gerekli fonksiyonun veya eşitliğin hesaplanması ve tahminlerin oluşturulması işlemi **regresyon analizi** olarak adlandırılır.

Regresyon analizinde değişkenler arasındaki ilişki doğrusal olabileceği gibi eğrisel de olabilir.

Eğer değişkenler arasındaki ilişki doğrusal ise doğrusal regresyon analizi, değilse doğrusal olmayan regresyon analizi adını alır.



# Basit Doğrusal Regresyon Analizi

Regresyon denklemi, bir bağımlı ve bir bağımsız değişkenden meydana geliyorsa ve değişkenler arasındaki ilişki doğrusal ise yapılan regresyon analizi basit doğrusal regresyon analizi adını alır.

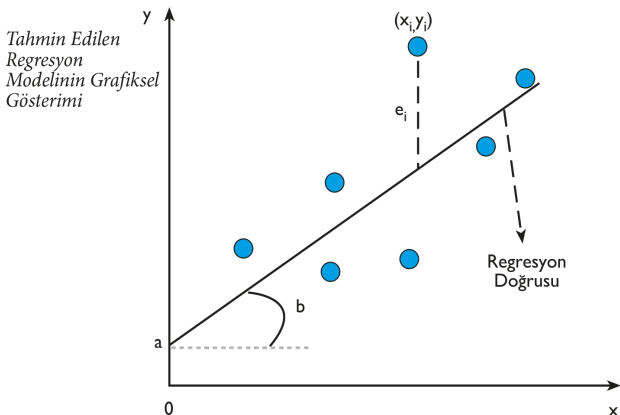
Bu ünite sadece değişkenler arasındaki doğrusal ilişki incelenecektir ve doğrusal modelin nasıl tahminleneceği konusu işlenecektir.

Doğrusal regresyon denkleminin tahmini için bir çok teknik kullanılmaktadır.

En yaygın kullanılan teknik olan en küçük kareler tekniği bu ünite içerisinde ele alınacaktır.



# Basit Doğrusal Regresyon Analizi



Tahmin edilen regresyon modelinin grafiksel gösterimidir. Burada modelin parametre tahminleri için en küçük kareler analizi kullanılmıştır. En küçük kareler analizi hataları en küçükleme tekniğidir.



# Korelasyon Analizi

$n$  hacimli örneklemden elde edilen veri setinde  $x$  ve  $y$  değişkenleriyle ilgili gözlem değerleri yer alır.  $x$  ve  $y$  değişkenleri arasında var olan kuramsal ilişki, izleyen doğrusal model yardımıyla araştırılabilir. Basit doğrusal regresyon modeli,

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

olarak yazılır. Model'de

- $y_i$ ; bağımlı değişken  $y$ 'nin  $i$ 'inci gözlem değerini,
- $\beta x_i$ ; bağımsız değişken  $x$ 'in  $i$ 'inci gözlem değerini,
- $\varepsilon_i$ ;  $i$ 'inci gözlem için ortalaması sıfır ve tüm gözlemler için sabit  $\sigma$  standart sapmalı normal dağılıma sahip olduğu varsayılan rassal hatayı,



# Korelasyon Analizi

$\alpha$  ve  $\beta$ ; tahminlenecek parametre değerlerini gösterir.  $\alpha$ ,  $x = 0$  olduğunda  $y$ 'nin alacağı değeri,  $\beta$ ,  $x$ 'te meydana gelecek birim değişikliğin  $y$ 'deki oransal değişmeyi göstermektedir. Ancak uygulamada her zaman evren değerlerinin tamamına ulaşmak mümkün olmadığından örnekleme başvurulur, örneklem için  $x$  ve  $y$  değişkenleri arasındaki kuramsal ilişki aşağıdaki doğrusal tahmin denklemi ile araştırılabilir:



# Korelasyon Analizi

$$y_i = a + bx_i + e_i$$

Bu modelde:

- $y_i$ ; bağımlı değişken  $y$ 'nin  $i$ 'inci gözlem değerini,
- $x_i$ ; bağımsız değişken  $x$ 'in  $i$ 'inci gözlem değerini,
- $e_i$ ;  $i$ 'inci gözlem için ortalaması sıfır ve tüm gözlemler için sabit standart sapmalı normal dağılıma sahip olduğu varsayılan rassal hatayı,

$a$  ve  $b$ ; sırasıyla  $a$  ve  $b$ 'nin tahmin değerlerini gösterir. Benzer şekilde  $a$ ,  $x = 0$  olduğunda  $y$ 'nin değerini,  $b$  ise  $x$ 'te meydana gelecek birim değişikliğin  $y$ 'deki oransal etkisini gösterir.



Ödev: Coğrafya için İstatistik kullanımına; tez, makale, çalışma, kitap vb. bir örnek bulunuz, en fazla bir sayfada açıklayınız.





Ödev: T-Testi ve ANOVA testi nedir?



## Kaynaklar I

- [1] K. Mert Çubukçu,  
*"Planlamada ve Coğrafyada Temel İstatistik ve Mekansal İstatistik"*,  
Ankara: Nobel Akademik Yayıncılık Eği. Dan. Tic. Ltd. Şti., (2015).
- [2] A. Özmen, F. Er, M. Atlas, E. Şıklar,  
*"İstatistik (AÖF)"*,  
Eskişehir: Anadolu Üniversitesi Yayınları, (2012).
- [3] L. İşbilen Yücel,  
*"İstatistik Maliye Uzaktan Eğitim"*,  
İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi Ders Notu.
- [4] Ö. Serper,  
*"Uygulamalı İstatistik 1"*,  
Bursa: Ezgi Kitapevi, (2004).



## Kaynaklar II

- [5] Murat Komisyon,  
*"İstatistik"*,  
Murat Açıköğretim Yayınları, (2004).
- [6] N. Gürsakal, A. Oğuzlar,  
*"Betimsel İstatistik"*,  
Dora Yayıncılık, (2019).
- [7] Y. Baykul, C. O. Güzeller,  
*"Sosyal Bilimler için İstatistik Uygulamaları"*,  
Ankara: Pegem Akademi, (2014).
- [8] Ankara Üniversitesi Açık Ders Sunumları,  
*"AKT102 İSTATİSTİK, BÖLÜM 3 OLASILIK"*,  
<https://acikders.ankara.edu.tr>

