

# IST2084/ IST104.1/ IST104.2

## Biyostatistik

- Doç. Dr. Fatih KIZILASLAN
- <http://mimoza.marmara.edu.tr/~fatih.kizilaslan/>

# İstatistik ?

Sayısal veri yığınlarının  
toplanması,  
analiz edilmesi,  
yorumlanması ve  
sunulması ile ilgilenen bir bilim dalı...

## Biyostatistik

İstatistiksel metodları biyoloji, tıp ve sağlık bilimlerine uygulayan bir bilim dalıdır.

### İstatistiğe neden ihtiyaç duyarız ?

- Değişkenlik ve çeşitlilik söz konusu olmasaydı istatistiksel yöntemlere gerek kalmazdı.
- Değişkenliği tanımlamak, analiz etmek ve sınıflandırmak istatistiğin temel konularıdır.

Toplam Değişkenlik=Açıklanabilen Değişkenlik+ Tesadüfü Değişkenlik

- Değişkenlik bir gözleme ait nitelik olmayıp ancak tekrarlı gözlemler sonucu ortaya çıkmaktadır.
- Her tür değişkenlik istatistiğin çalışma konusu değildir.
- İstatistik **rassallık** içeren değişkenlikle ilgilenir. (van Belle, 2002)
- **Rastgelelik:** Tesadüfilik, rassallık (randomness)
- **İstatistik** rastgelelik ortamında analiz ve sonuç elde edebilmemizi sağlar.
- **Matematik** ise sebep sonuç ilişkileri kesin olan durumlarda kullanılabilir.

## Matematiksel Modeller

- Deterministik modeller: Sebep sonuç ilişkileri kesin olan modeller.
- Rassal (Stokastik) Modeller: Rassallık içeren modeller.

Rassallık içeren modelleri çözümlene İstatistiğin konusudur.

**Peki bu çözümlene nasıl yapılır?**

**Olasılık Teorisi** deney yapmadan yığın olaylarının uyduğu kuralları belirler.

Modellerden yola çıkarak rassal olaylar yığınının uyduğu kuralları belirler.

- İstatistik ise veri bilimidir.
- Verilerden yola çıkarak uygun olasılık modeli belirlenir.
- Belirlenen model için olasılık teorisinin kuralları kullanılarak sonuçlar elde edilir.
  
- **Olasılıkta tüm anakütle bilinir.** Olasılık kuralları ile anakütle ve onun alt birimleri için kurallar belirlenir.
  
- **İstatistikte ise tüm anakütle mevcut değildir.** Anakütlenin alt birimleri kullanılarak uygun bir olasılık modeli belirlenerek tüm anakütle hakkında bilgiler ve sonuçlar elde edilir.

# 1. Bölüm: Temel Kavramlar

- **Kitle (popülasyon anakütle, anakitle, yığın, evren, population)**

İstatistikte ilgilenilen özelliğe sahip bütün birimlerin oluşturduğu topluluk.

Örneğin Türkiye'de herhangi bir X takımının tüm taraftarlarının oluşturduğu topluluk.

- **Örneklem (örnek, sample)**

Kitleden rastgele seçilen daha az sayıda birimden oluşan topluluk.  
Kitlenin alt kümesidir.

Örneğin X takımının taraftarlarında rastgele seçilen 1000 kişilik topluluk.

- **Parametre**

Anakitle hakkında bilgi veren ortalama, standart sapma gibi karakteristik deęerler parametre olarak adlandırılır.

- **Deęişken (variable)**

İlgilenilen kitlenin bir karakteristięi veya özellięidir.

Birimlerin farklı deęerler alabildikleri nitelik veya niceliklerdir.

Örneęin kişilerin göz renkleri, medeni durumları, yaşları, boyları, kiloları gibi.



- Bir deęişken sayısal deęerlerle ölçülebiliyorsa bu deęişkene **nicel (sayısal) deęişken** denir. Örneęin aęırlık, uzunluk, sınavdan alınan not, hasta sayısı gibi.
- Nicel deęişkenler aldıkları deęerlere **Kesikli** ve **Sürekli** olmak üzere **ikiye** ayrılır.

**Kesikli deęişken:** Sadece sonlu (sınırlı) sayıda deęerler alabilen deęişkenlerdir. Örneęin bir markete gelen müşteri sayısı, trafik kazası sayısı, otoparktaki araç sayısı, sınıftaki öğrencilerin sayısı gibi.

**Sürekli deęişken:** Sonsuz sayıda deęerler alabilen deęişkenlerdir.

Örneęin kişilerin aęırlıkları, uzunlukları, bir elektronik cihazı ömür süresi gibi.

- Bir deęişken sayısal deęerlerle ölçülemiyorsa bu tip deęişkenler **nitel (kategorik) deęişken** olarak adlandırılır.

Örneęin cinsiyet, medeni durum, göz rengi, saç rengi gibi.

- Araştırmalarda önemli konularda biri ilgilenilen deęişkenin **nasıl ölçüleceęidir**.
- **Ölçme** herhangi bir nitelięi gözlemek ve gözlem sonucunu sayılarla veya başka sembollerle ifade etmektir.
- Deęişkenlerin ölçümleri genel olarak dört farklı biçimde yapılabilir.

# Ölçek (Ölçüm) Türleri

- **Sınıflama Ölçeği (Nominal Ölçek):**

Gözlem sonuçları benzerliklerine/farklılıklarına göre sınıflandırılıyorsa bu ölçek sınıflama ölçeğidir. Örneğin, cinsiyet (kadın ve erkek), meslek (öğretmen, mühendis, doktor..) gibi.

Bu ölçekte ölçülmüş veriler üzerinde hiçbir aritmetik işlem yapılmaz, anlamı yoktur. Ancak sınıfların frekanslarına dayanan istatistiksel işlemler yapılabilir.

## • Sıralama Ölçeği (Ordinal Ölçek):

Gözlem sonuçları bir sınıflamaya tabi tutulmakla beraber belli bir özelliğe sahip olma bakımından sıralanabiliyorsa bu sıralama ölçeğidir.

Örneğin:

Bir yarışmanın sonucundaki dereceler: Birinci, İkinci, Üçüncü ...

Bir kişinin sağlık durumu: kötü, orta, iyi, çok iyi gibi.

Eğitin durumu: İlköğretim, Lise, Lisans, Y. Lisans gibi.

Bu sınıflar (kategoriler) arasında büyüklük, küçüklük sıralamaları yapılabilir. Sınıflar için kullanılan harfler alfabetik olarak, sayılar ise büyüklüklerine göre sıralanmış olur.

Örneğin bir yemek hakkındaki beğeninizi kötüden iyiye olmak üzere 1,2,3,4 ile belirtelim. Bu durumda 1 ile 2 arasında beğeni açısından 2 kat fark yoktur. **Bu sayısal değerler matematiksel olarak bir anlam taşımaz.**

- **Eşit Aralıklı Ölçek (Interval Ölçek):**

Gözlem sonuçlarını bir özelliğe sahip olma miktarı bakımından sıralamakla beraber bunlar arasındaki uzaklığı da gösterme imkanı veren ölçme türüdür. Bu ölçekte ölçüm boyutunun tüm değerler arasında uzaklık aynıdır.

Bu ölçekte başlangıç noktası keyfi olarak seçilir.

Sıfır değeri ölçülen özelliğin geçekten bulunmadığı anlamına gelmez.

Örneğin termometre gördüğümüz 0 santigrat (C) bir anlamı vardır.

Örneğin, Antalya (24 C), Ankara (19 C) ve Kars (11 C) şehirlerinin yıllık ortalama sıcaklıklarına göre eşit aralıklı ölçme düzeyinde sıcaklıkları santigrat birimi ile ölçeriz.

<u>Şehir</u>	<u>Sıralama Ölçeği</u>	<u>Eşit Aralıklı Ölçme</u>
Antalya	1 (10)	24 C
Ankara	2 (20)	19 C
Kars	3 (30)	11 C

Sıralama ölçeğinde kullanılan 1,2,3 sadece sırayı gösterir (10,20,30 da kullanabiliriz).

Eşit aralık ölçeğinde ölçüm değerleri arasındaki farklar örneğin  $24-19=5$  ile  $19-11=8$  değerleri Antalya ile Ankara arasındaki sıcaklık farkının Ankara ile Kars arasındakinden az olduğunu gösterir.

## • Oran Ölçeđi:

Eđit aralıklı ölçme düzeyinde yapılan ölçme mutlak bir sıfır (başlangıç) noktasına göre yapılabiliyorsa bu ölçme düzeyine oranlama ölçme düzeyi denir.

Örneđin ađırlık, hız, uzunluk, aylık gelir gibi deđişkenler için mutlak sıfır noktası oran ölçeđi ile ölçülür. Bunlar için başlangıç noktası 0 (sıfır) dır.

Örneđin 0 kg, 0 cm ... yokluđu ifade eder.

Ölçümler arasında oranlama yapılabilir. İki kişinin ađırlıkları 50 kg ve 100 kg olsun. Bu durumda 100 kg olan kiři 50 kg olana göre 1 kat daha ađırdır.

- Sınıflama ve sıralama ölçeklerinde ölçülen değişkenler **nitel (kategorik-sayısal olmayan) değişkenlerdir.**
- Eşit aralık ve oran ölçeklerinde ölçülen değişkenler **nicel (sayısal) değişkenlerdir.**



- İstatistiğin kullanımını iki yolla olur.

## 1. Betimleyici (tanımlayıcı, **descriptive**) istatistik

Elde edilen verilerin sınıflandırılması, sıklık dağılımlarının oluşturulması, merkezi eğilim ölçütlerinin (aritmetik ortalama, ortanca, tepe değeri vb.) ve dağılım ölçütlerinin (varyans, standart sapma, standart hata, değişim aralığı, yüzdeler vb.) hesaplanması, bulguların tablo ve grafikler yardımıyla sunumu betimleyici istatistiğin konularını oluşturur.

## 2. Çıkarımsal (**inferential**) istatistik

Örnekten elde edilen bulgularla örneğin elde edildiği anakitle hakkında tahminlerde bulunma, karşılaştırmalar yapma ve kararlara varma işlemleri çıkarımsal istatistiğin alanına girmektedir.

## 2. Bölüm: Verilerin Düzenlenmesi

- **Frekans dağılımları ve grafikler**

Kitle veya örneklemden belirli değişkenler için çeşitli ölçüklere göre alınan ölçüm (gözlem) sonuçlarının oluşturduğu kümeye **veri** denir.

Verinin özetlenmesi (tanımlanması) için yapılan işlemlerden biri frekans dağılımlarını oluşturmaktır.

Nitel veriler için kategorik olarak frekans dağılımı yapılır. Her bir kategoride (örneğin meslek grupları) bulunan birim sayısına göre oluşturulur.

- Örnek: 40 ailenin çocuk sayılarına göre frekans dağılımı aşağıdaki gibidir. Çocuk sayısı kesikli değişkendir.

<u>Çocuk sayısı</u>	<u>Aile sayısı (frekans)</u>
0	8
1	13
2	14
3	2
4	2
5	1

Sürekli değişken için değer aralıkları alt aralıklara yani **sınıflara** bölünür. Bu sınıfların kesişimleri boş küme olmalıdır. Sınıf sayısının kaç olacağı belirlenmelidir.

Eğer sınıf sayısının ne olacağı belirtilmemiş ise Sturges formülü ile sınıf sayısı ( $k$ )

$$k = 1 + 3.3 \log(n)$$

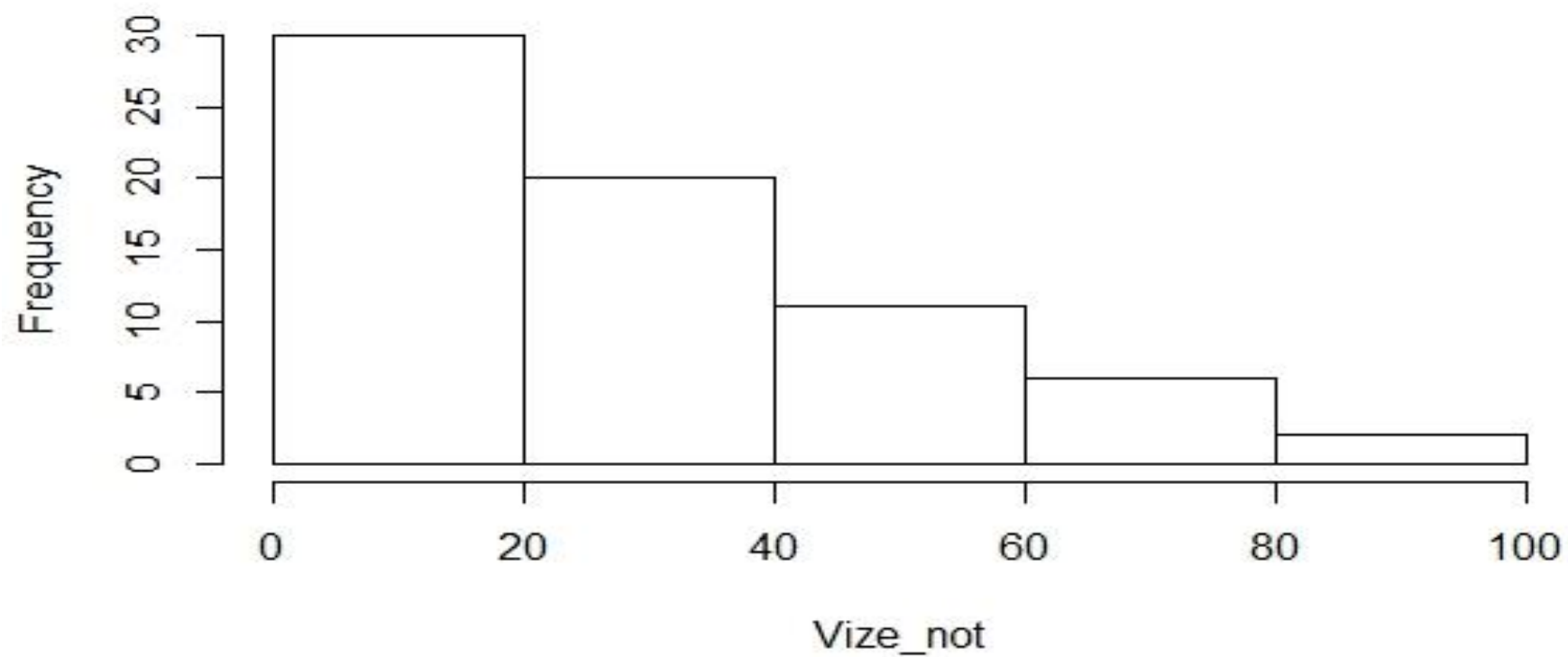
biçiminde belirlenir ( $n$  gözlem sayısı).

- Sınıf aralığı (genişliği): İki sınıf arasındaki farktır.

$$\text{sınıf aralığı} = \frac{\text{en büyük değer} - \text{en küçük değer}}{k}$$

**Histogram:** Koordinat eksenleri üzerinde her sınıf için çizilen dikdörtgenlerden oluşan grafikdir. Bu dikdörtgenlerin taban kenar uzunlukları sınıf aralığına eşittir. Diğer kenar uzunluğu da sınıfların frekansına eşittir.

**Histogram of Vize\_not**



**Örnek 1:** 30 farklı firmada çalışan sayıları aşağıda verilmiştir.

26, 29, 10, 30, 31, 46, 17, 35, 18, 16, 24, 30, 32, 35, 38,  
24, 27, 25, 36, 30, 17, 23, 18, 18, 44, 41, 18, 24, 36, 38

**Örnek 2:** Bir kütüphanede 30 öğrencinin bir dönem boyunca geçirdikleri süreler (saat) aşağıda verilmiştir. (Örnek 1 de verilen aynı veri)

26, 29, 10, 30, 31, 46, 17, 35, 18, 16, 24, 30, 32, 35, 38,  
24, 27, 25, 36, 30, 17, 23, 18, 18, 44, 41, 18, 24, 36, 38