

# AST206 ASTROİSTATİSTİK

## 1. KONU

Hazırlayan: Doç. Dr. Tolgahan KILIÇOĞLU

### 1. İSTATİSTİKTE TEMEL KAVRAMLAR

#### 1.1 İstatistik Nedir?

Bilimsel bir çalışma yaparken çalışmanızın belirli noktalarında bazı kararlar vermek zorunda kalırsınız. Örneğin, astronomide bir konuda cevabını bulmak istediğiniz bir soru var ve bu konu ile ilgili gözlemevinde bir gözlem yaptınız. Bu noktada kendinize şu soruları sormanız gerekecektir: Elde ettiğim gözlem verileri cevap aradığım soruya ışık tutacak kadar duyarlı mı? Veriler kendi içinde ve başka yollarla elde edilen verilerle tutarlı mı? Sadece bu gözlem kullanılarak bu soru cevaplanabilir mi yoksa daha fazla gözlem yapmam gerekli mi? Eğer veri yeterliyse veriyi analiz ettikten sonra ilgilendiğim soruya hangi doğrulukta cevap verebileceğim? Eğer bu soru için öne sürdüğüm bir hipotez varsa bu hipotezin doğruluğunu nasıl test edebilirim? Bu ve buna benzer sorulara cevap verebilmenin tek bir güvenilir anahtarı bulunmaktadır: **İstatistik!**

İstatistik bir verinin;

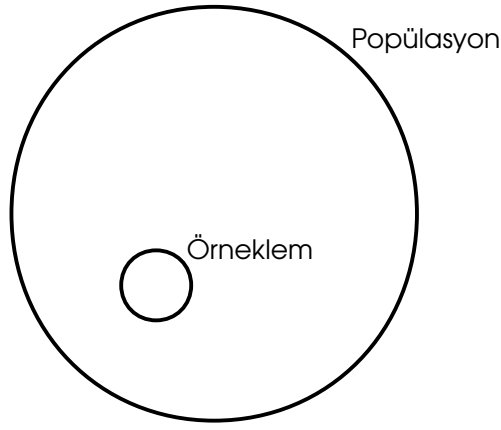
- i) toplanmasını,
- ii) sınıflandırılmasını,
- iii) sunulmasını,
- iv) çözümlenmesini ve
- v) yorumlanmasını

bilimsel ve teknik yöntemlerle gerçekleştirmeyi konu alan bir bilim dalıdır.

#### 1.2 Popülasyon ve Örneklem

Üzerinde istatistiksel çalışma yapılması istenen ve belirli bir özelliğe sahip olguların tamamına popülasyon denir. Örneğin, "Ankara'da yaşayan 30 yaşını doldurmuş avukatlar" dendiğinde bu tanıma giren tüm insanlar bir popülasyon oluşturur. Benzer şekilde, "değen çift yıldızlar" üzerine bir istatistiksel çalışma yapılacaksa evrendeki tüm değen çift yıldızlar bu çalışmanın popülasyonunu oluşturacaktır.

Örneklem ise bir popülasyon içerisinde seçilen ve o popülasyonun tamamını temsil edebilecek özelliklere sahip olduğu düşünülen bir alt gruptur (Şekil 1.1). Örneğin evrende gözlenen değen çift yıldızlardan rastgele seçilen yeterli sayıda cisim yukarıda verilen popülasyon için bir örneklem grubu olarak kabul edilebilir. Ancak, sadece açık yıldız kümelerinde bulunan değen çift yıldızlar seçilirse bu gruba örneklem demek doğru olmaz. Çünkü bu grup popülasyonun tamamını temsil etmemektedir. Bir örneklemin birimleri mutlaka **rastgele** seçilmelidir.



**Şekil 1.1** Popülasyon ve örneklem

### 1.3 Veri ve Veri Türleri

Bir popülasyonun veya örneklemin gözlemlerinden (veya onlara yapılan bir anket veya deneyden) veriler elde edilebilir. Toplanan veriler temelde üç farklı türe sahiptir: Nitel, Sıralama ve Nicel.

#### 1.3.1 Nitel veri

Nitel veri evet/hayır gibi ifadelerden, E/H gibi harflerden veya kadın/erkek gibi kategorilerden oluşur. Nitel veri bazen sayı da içerebilir; ancak bu sayıların üzerinde matematiksel bir işlem yapmak anlamsızdır. Bunlara örnek olarak posta kodu, telefon numarası veya öğrenci numarası verilebilir. Çizelge 1.1’de Pleiades açık yıldız kümesine üye yıldızların çift olup olmadıklarına ilişkin nitel bir bilgi verilmektedir. Evet (E) cevabı yıldızın çift olduğunu gösterirken hayır (H) çift olmadığını gösterir. B harfi ise çift olma durumunun belirsiz/bilinmiyor olduğunu ifade etmektedir.

**Çizelge 1.1** Nitel veri: Pleiades kümesi üyesi yıldızlardan seçilen bir örneklemin çift olma durumu

E	H	E	H	B	E	E	E	B	E
H	B	E	B	H	B	E	B	E	E
E	H	E	E	E	E	H	E	H	B
B	E	H	B	E	E	E	H	E	H
E	E	B	H	E	H	E	E	B	E
E	E	E	E	H	H	E	H	B	E
H	B	E	H	E	B	E	E	H	E
B	H	E	E	E	E	E	H	E	E
E	B	H	E	E	E	E	B	E	H
E	H	E	B	B	H	E	E	H	...

### 1.3.2 Sıralama Verisi

Bir ölçüte dayanarak örneklem sıralanır ve örneklem üyelerine bir sıra numarası atanırsa bu veriye **sıralama verisi** adı verilir. Örneğin bir araba yarışının sonucunda yarışmacıların kaçınıcı oldukları bir **sıralama verisi**dir. Tablo 1.2'de gökyüzündeki yıldızlar en parlaktan en sönüğe doğru sıralanmıştır. Bu yıldızlara atanan sıra numaraları yine bir sıralama verisidir. Sıralama verileri üzerinde matematiksel işlemler yapmak yine anlamsızdır. Renk sırası yapmak da bir sıralama verisi oluşturabilir. Renkler arasında matematiksel bir işlem yapılmasının yine bir anlamı olmaz.

**Çizelge 1.2** Sıralama verisi: Gökyüzündeki yıldızların görünür görsel parlaklıklarına göre sıralanması

Sıra No.	Yıldız	Sıra No.	Yıldız
1	Güneş	6	Vega
2	Sirius	7	Capella
3	Canopus	8	Rigel
4	Rigel Kentaurus	9	Procyon
5	Arcturus	...	...

### 1.3.3 Nicel Veri

Numaralardan oluşan ve bir ölçümü veya sayımı veren veri türüne **nicel veri** adı verilir. Nicel veriler üzerinde matematiksel işlemler yapılması anlamlıdır. Örneğin boy, kilo, maaş, sıcaklık, hız, kütle, üye sayısı, zaman vb. kavramlar nicel verilere örnektir. Çizelge 1.3'te nicel veriye örnek olarak Güneş sistemindeki gezegenlerin kütleleri verilmektedir.

**Çizelge 1.3** Nicel veri: Güneş Sistemi'ndeki gezegenlerin kütleleri

Gezegen	Kütle	Gezegen	Kütle
Merkür	$3.301 \cdot 10^{23}$ kg	Jüpiter	$1.899 \cdot 10^{27}$ kg
Venüs	$4.867 \cdot 10^{24}$ kg	Satürn	$5.685 \cdot 10^{26}$ kg
Yer	$5.972 \cdot 10^{24}$ kg	Uranüs	$8.682 \cdot 10^{25}$ kg
Mars	$6.417 \cdot 10^{23}$ kg	Neptün	$1.024 \cdot 10^{26}$ kg

Nicel verinin iki alt türü bulunur: Kesikli ve Sürekli veri.

**Kesikli nicel veri:** Bu veri türü bir sayım gibi düşünülebilir. Örneğin bir fabrikanın belirli bir günde ürettiği dizüstü bilgisayar sayısı veya bir yurttaki öğrencilerinden Instagram kullananların sayısı kesikli veriye örnektir. Her iki örnekte de verilen sayılar tamsayı olmak zorundadır. Kesikli verinin bir diğer özelliği alabileceği değerlerin miktarının da sonlu olabilmesidir. Örneğin bir zar düşünelim. Standart bir zar atıldığı zaman sadece 1, 2, 3, 4, 5, ve 6 değerlerini verebilir. Şimdi zarın kenarlarının üzerine tamsayılar yerine kesirli sayılar

yazalım: örn., 1.548, 5.879, 7.652, 2.471, 3.743, ve 0.256. Böyle bir zarı attığımızda gerçekten de zar bize küsüratlı sayılar verecektir; ancak yine de elde edilen veri kesikli olacaktır. Çünkü zar bu sayıların arasında bir değer verememektedir. Verebileceği değerlerin miktarı yine altı adettir.

**Sürekli nicel veri:** Genellikle bir ölçüm sonucunda elde edilen ve sonsuz sayıda mümkün değeri olabilen verilerdir. Sıcaklık, kadir, kütle, yarıçap vb. sürekli veriye örnektir. Çizelge 1.3'de sunulan nicel veriler (kütle) de bir sürekli veri örneğidir.

## 1.4 Ölçüm Seviyeleri

Soruların eşit puanlı olduğu bir sınavdan 80 alan öğrenci ile 40 alan öğrenci karşılaştırılırken bu iki değer oranlanarak birinci öğrencinin ikinci öğrenciden sınavda iki kat daha başarılı olduğu söylenebilir. İki not arasındaki farktan da birinci öğrencinin kaç soru daha fazla bulunabilir. Ancak sayılar her zaman bir miktar ifade etmeyebilir. Örneğin, bu iki öğrencinin TC kimlik numaraları arasındaki farkın veya oranın pratikte hiçbir anlamı yoktur. Bu nedenle bir istatistik çalışma yaparken her ölçüme aynı şekilde davranılamaz. Hangi aritmetik işlemin veya istatistiksel sürecin hangi tür ölçüme uygulanması gerektiğinin bilinmesi gerekir. Bu noktada ölçüm seviyeleri göz önünde bulundurulur.

### 1.4.1 Nominal Ölçüm

Nominal ölçümler sıralı olmayan nitel verilerdir. Örnek olarak bir anketin cinsiyet bölümündeki kadın ve erkek seçeneği verilebilir. Bu iki grubun belirli bir sırada verilmesi söz konusu değildir. Nominal ölçümler sayılarla da ifade edilebilir. Örneğin et yiyenlerin, vejetaryenlerin ve veganların sırasıyla 1, 2, 3 numaralarıyla temsil edildiği bir anket düşünelim. Buradaki sayılar bir sıralama değildir ve değerlerinin hiçbir anlamı yoktur (veganların numarasının et yiyenlerinkinin üç katı olması bir anlam ifade etmez). Galaksi'nin diskinde bulunan yıldızların  $D$ , halosunda bulunan yıldızların ise  $H$  olarak işaretlendiği bir astronomi kataloğu da benzer şekilde nominal bir ölçüm içerir. Nominal ölçümlere ilişkin önemli kurallardan biri bu ölçümlerin ortalamasının alınamamasıdır. Alındığı taktirde elde edilen değer bir anlam ifade etmez. Bu kısıtlamalar nedeniyle nominal veriler üzerinde sınırlı sayıda istatistiksel çalışma yapılabilir.

### 1.4.2 Ordinal Ölçüm

Ordinal ölçümler sıralama verileridir. Örneğin bir sınavda öğrencilerin birinci, ikinci veya üçüncü olarak adlandırılması bir ordinal ölçümdür. Ordinal ölçümlerin en önemli özelliği sıra numaraları arasındaki matematiksel farkın fiziksel bir anlam ifade etmemesidir: bir araba yarışında birinci olan kişinin ikinci kişiden daha hızlı gittiğini söyleyebiliriz; ancak ne kadar hızlı gittiğini söyleyemeyiz. Benzer olarak, yıldızları parlaklıklarına göre sıraladığımız Çizelge 1.2'de 1 numaralı yıldızın 2 numaradan daha parlak olduğunu söyleyebiliriz; ancak ne kadar parlak olduğunu hakkında bir yorumda bulunamayız. Verilen tüm bu örneklerde 1. ile 2. arasındaki farkla 2. ile 3. arasındaki farkın birbirlerine eşit olması gerekmez. Yani sıralamadaki numaralar eşit aralıklı değildir. Bu nedenle ordinal ölçümler aritmetik işlemler yapmak uygun veriler değildir. Ancak yine de bu verilerin işlenebilmesi için birkaç istatistiksel metod bulunmaktadır.

### 1.4.3 Aralık Ölçümü

Ordinal ölçümde olduğu gibi aralık ölçümü de sıralı bir ölçümdür. Ancak burada ölçümler arasındaki farklar birbirlerine eşit ve anlamlıdır (örn. Sıcaklık). Aralık ölçümlerinin **doğal sıfır noktası bulunmaz**. Örneğin, hava sıcaklığının 0 °C olması havada hiçbir enerjinin olmadığı anlamına gelmez. Aralık ölçümleri arasında fark alma işlemleri yapılabilirken birbirleriyle çarpma/bölme işlemleri yapılamaz. 40 °C sıcaklığındaki bir hava 20 °C sıcaklığındaki bir havadan iki kat (yani 40/20) daha sıcak **değildir** (eğer öyle olsaydı, 10 °C sıcaklığındaki havanın da 0 °C havadan sonsuz kez daha sıcak olması gerekirdi).

Aralık ölçümlerinde toplama/çıkarma işlemleri yapılabildiğinden ortalama da alınabilmektedir. Örneğin 40 °C ile 20 °C nin ortalaması 30 °C dir. Bu sıcaklıkların başka bir ölçekteki (örn. Fahrenheit) değerleri kullanılarak da ortalama alınsa bulunan sonuç °C ye çevrildiğinde yine 30 °C olduğu görülecektir.

Aralık ölçümüne verilebilecek en güzel örneklerden biri gökcisimlerinin koordinatlarıdır. Bir gökcisiminin dikaçıklığının 0° olması bir yokluğu değil sadece önceden kabul edilmiş bir doğrultuyu (ekvator düzlemini) işaret etmektedir. İki yıldızın dikaçıklıkları arasındaki fark onların birbirlerine olan açısal uzaklıklarına ilişkin bir bilgi verirken, dikaçıklıklarının çarpılması veya bölünmesi bizi herhangi bir anlamlı bilgiye götürmez.

### 1.4.4 Oranlanabilir Ölçüm

*Oranlanabilir ölçümün aralık ölçümünden farkı **doğal sıfır noktasının** olmasıdır.* Örneğin boy ve kilo ölçümleri oranlanabilir ölçeklerdir. Gerçekten de 120 kg olan birisinin 60 kg olandan iki kat daha kütleli, veya boyu 120 cm olan birisinin 80 cm olandan 3/2 kat daha uzun olduğu rahatlıkla söylenebilir. 0 kg ve 0 cm ifadeleri sırasıyla bir kütlenin ve bir uzunluğun olmadığını ifade eden doğal sıfır noktalarıdır. Bir para miktarını ifade etmek de oranlanabilir ölçüm türüne girer. Astronomide gök cisimlerinin kütleleri, yarıçapları, yoğunlukları, uzaklıkları, paralaksarı, hızları vb. doğal sıfıra sahip olan oranlanabilir ölçeklerdir.

*Aralık ölçümlerinin bazıları oranlanabilir ölçümlere dönüştürülebilirler.* Örneğin, sıcaklık parametresine geri dönülürse, mutlak sıfır sıcaklık Celcius ölçeğinde -273.15 °C'e karşılık gelmektedir. Bu nedenle Celcius ölçeğinde verilen sıcaklıklardan 273.15 sayısı çıkarılırsa (yani Kelvin'e dönüştürülürse) ölçek mutlak sıfıra (doğal sıfıra) sahip olacağından hem aralıklı hem de oranlanabilir olur. Bilimsel çalışmalarda genellikle Kelvin ölçeğinin tercih edilmesinin sebebi budur.

Oranlanabilir ölçümler istatistikte en kullanışlı ölçümlerdir. Bu ölçümler üzerinde toplama/çıkarma/çarpma/bölme işlemleri yapılabilir, ortalama alınabilir ve birçok istatistiksel süreç uygulanabilir.

## 1.5 Deney ve Gözlem

Deney ve gözlem birbirlerinden farklı kavramlardır. Şimdi bu iki kavram arasındaki farklılığa ve bu farklılıktan ileri gelen bazı kavramlara göz atalım.

Deneyde araştırmacı özel bir etkinin kime veya neye uygulanacağını kendisi seçer. Belirli bir amaçla araştırmacı tarafından ayarlanan değişkene **bağımsız değişken** adı verilir. Eğer başka bir değişken bu bağımsız değişkenin değerinden bir şekilde etkileniyorsa bu değişkene **bağımlı değişken** adı verilir. Bağımlı değişken deney esnasında veya sonunda ölçülen değişkendir. Örneğin, bir araştırmacı yüksek sesli müziğin ders çalışma üzerindeki etkisini ölçmek istediğinde yüksek, orta güçlü ve düşük seviyede müzik sesi içeren üç farklı odaya belirli sayıda denekler yerleştirir ve bir ders notunu çalışmalarını ister. Daha sonra denekler okudukları ders notuna ilişkin bir sınava tabi tutulur. Burada müzik sesi bağımsız değişken, sınav notları ise bağımlı değişkendir. Araştırmada "müzik sesinin sınav notları üzerinde etkili olduğu" sonucuna varılabilir. Ancak bunun tersinin olması mümkün değildir. Yani, değişkenlerin yerini değiştirdiğimizde "sınav notları müzik sesinin seviyesini üzerinde etkilidir" cümlesi görüldüğü gibi hiçbir anlam ifade etmez. Bağımlı ve bağımsız değişkenler bu nedenle deneylerde net bir şekilde birbirlerinden ayrılırlar.

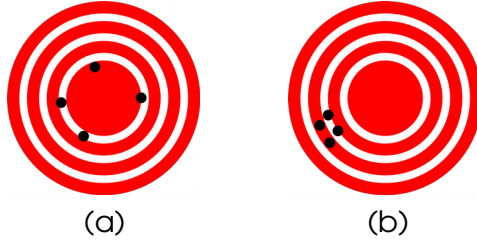
Gözlemde ise durum daha farklıdır. Bir gözlem yapıldığında gözlemcinin değişkenler üzerinde değişiklik yapması mümkün değildir. Örneğin, evrende Cepheid olarak adlandırılan değişen yıldızlar mevcuttur. Bir gözlemci bu türden yıldızları gözlediğinde değişim dönemleri ile mutlak parlaklıkları arasında sıkı bir ilişkinin olduğunu görür. Peki bunlardan hangisi bağımsız hangisi bağımlı değişkendir? Yani yıldızın dönemi değiştiği için mi parlaklığı değişir yoksa parlaklığı değiştiği için mi dönemi değişir? Görüldüğü gibi bir gözlemde değişkenlerin hangisinin bağımsız ve hangisinin bağımlı olduğu belirsiz olabilir. Bu örnekte gerçekte dönem ve parlaklık parametrelerinin her ikisi de bağımlı değişkenlerdir ve başka bir üçüncü değişkene bağılırlar: yıldızın kütlesi. Bu nedenle bir gözlem sonucunda elde edilen değişkenler araştırmanın başında bağımlı veya bağımsız değişkenler olarak ayrılmaz ve **gözlemsel değişkenler** olarak adlandırılırlar.

## 1.6 İstatistiğin Amacı

İstatistiğin amacı iki temelde özetlenebilir;

- i) Bir popülasyondan/örneklemden toplanan verinin işlenerek tablo ve şekiller yardımıyla sunulması (tanımlayıcı istatistik),
- ii) Bir örneklemden alınan parametrelerin olasılık kuramlarına uygun şekilde işlenerek popülasyona ilişkin parametrelerin tahmin edilmesi (çıkarımsal istatistik),

Yıllara göre üniversite tercihlerinde Astronomi ve Uzay Bilimleri bölümünü tercih eden öğrenci sayılarını gösteren bir grafiğe baktığınızda bir **tanımlayıcı istatistik** örneği görmüş olursunuz. Tanımlayıcı istatistik aynı zamanda yapılan bir gözlemden/deneyden elde edilen verilerin ne kadar duyarlı olduğunu (eğer gerçek değer biliniyorsa ne kadar doğru olduğunu) da ortaya koyar (bkz. Şekil 1.1). Böylelikle yapılan bir gözlemin sonuçlarının kullanılabilir olup olmadığı, kullanıldığı taktirde ne kadar hata ile bize bir bilgi sunabileceği ortaya konur.



**Şekil 1.1** (a) Yüksek doğruluk ama düşük duyarlılık, (b) yüksek duyarlılık ama düşük doğruluk

Belirli konular üzerine belirlenen örnekleme yapılan anketler ve bu anketlerin sonuçlarının yorumlanması ise **çıkarımsal istatistik** tir. Örneğin düzenli uyku uyumayanların daha fazla baş ağrısı çektiğini öne süren bir bilimsel çalışma bir çıkarımsal istatistiğin sonucundan elde edilir.

Astronomide çıkarımsal istatistik oldukça fazla kullanılır. Çünkü astronomide bir popülasyonun tamamının gözlenmesi çoğu zaman mümkün olmaz. Belirli bir tür gök cismine ait bir parametre inceleniyor olsun (örneğin, beyaz cücelerin yarıçapları). Eğer bu türden yıldızlar yeterli sayıda ve rastgele olacak şekilde gözlenir ve bir örneklem grubu oluşturulursa tüm beyaz cücelerin yarıçaplarına (popülasyona) ilişkin tahmini genellemeler yapılabilir. Örneğin, bu yarıçapların hangi aralıkta değiştiği, ortalamasının ne olduğu veya belirli bir yarıçap değerinde bir yığılma olup olmadığı ortaya konabilir. Yarıçaplarla birlikte kütleler de gözlenirse beyaz cüceler için bu iki parametre arasında bir ilişki olup olmadığı da ortaya konulabilir.