

TEMEL BİLGİLER

İSTATİSTİĞİN TANIMI

İstatistik denilince, belli bir olaya ait rakamlar topluluğu, yani *rakamsal veriler*, anlaşılır. Fakat bu rakamlar bir gözlem sonucu elde edilen rakamlardır. Örneğin milli gelir istatistikleri, dış ticaret istatistikleri, bir bölgedeki ailelerin aylık gelir düzeyleri, bir sınıftaki öğrencilerin boy ölçüleri vb. birer istatistik rakamlarıdır. Bu bakımdan, mesela logaritma ya da çarpım tablosunu oluşturan rakamlar istatistik sayılmazlar. Tekrar başa dönersek, istatistik rakamları belli bir olayın gözlenmesi ve ölçülmesi sürecinde ortaya çıkarlar. O halde istatistik rakamlarının ilgili olduğu bir olay olmalı, bu olay gözlemlenebilmeli ve ölçülebilmelidir.

Bir bilim dalı olarak İstatistik kelimesi, yukarıda sözü edilen gözlem ve ölçme sonucu elde edilen rakamsal verilerin bilgilerin tablolar veya grafikler şeklinde kullanıma sunulması ve bunların veri olarak karar alma sürecinde nasıl kullanılacağı hakkındaki yöntemlerin bilimi, yani araştırma metodolojisi anlamına gelmektedir. Bir bilim dalı olarak istatistiği şu şekilde tarif edebiliriz: *Belli bir konu ile ilgili verileri derlemek, düzenlemek, özetlemek, sunmak ve analiz ederek bu verilerden bir sonuca varmak için kullanılan yöntemler bütününe İstatistik adı verilir.* İstatistiğin üçüncü bir anlamı da *tahmin* dir. Bir veri kümesinden faydalanarak her hangi bir bilinmeyen parametre için tahmin olarak elde edilen yaklaşık değere de *istatistik* denilir. Böylece *istatistiğin*, veri, bilim dalı ve tahmin olmak üzere üç değişik anlamda kullanıldığını görmekteyiz.

İSTATİSTİĞİN KAPSAMI

İstatistik genel olarak iki kısma ayrılmaktadır. (1) Betimsel İstatistik (Deskriptif İstatistik, Tasviri İstatistik), (2) Tümevarımsal İstatistik (Çıkarsamalı İstatistik, Vardamlı İstatistik). Bir olaya ait gözlem ya da ölçme sonucu elde edilen verileri, bunları kullananlara tablo halinde, grafik halinde, yaklaşık tek bir değer (ortalama) halinde sunmak şeklindeki mevcut bir durumu tasvir etmeye yönelik istatistiksel yöntemlerin bütününe *Betimsel (Tasviri) İstatistik* adı verilir.

Mevcut duruma ait bilgilerden hareketle bir takım tahminlerde bulunma, yorumlarda bulunma, ya da bir örnekten hareketle yığın hakkında bazı çıkarsamalar yapmak amacıyla geliştirilen İstatistiksel yöntemlere de *Tümevarımsal İstatistik* adı verilmektedir. Tümevarımsal İstatistiğin altyapısını olasılık teorisi oluşturmaktadır. Tümevarımsal İstatistik adından da anlaşılacağı gibi özelden (örnekten,örneklem) genele (yığma, kitle) doğru bir prosedür takip etmektedir.

İstatistiğin, Teorik İstatistik ve Uygulamalı İstatistik şeklinde bir ayırıma tabi tutulduğu da görülmektedir. Bu ayırım daha çok istatistiğin öğretimi ile ilgilidir. İstatistiksel yöntemlerin nasıl geliştirildiği hangi matematiksel gerçeklere dayandığının izahı teorik istatistiğin konusudur. Geliştirilmiş istatistiksel yöntemlerin gerçek olaylara uygulanması ve bu esnada karşılaşılan sorunlar uygulamalı istatistiğin konusudur.

TEMEL KAVRAMLAR

Olay: İstatistiğin tanımı verilirken istatistiksel verilerin bir olaya dayanması gereğinden söz edildi. İstatistik açısından olaylar, Tipik olaylar ve Kollektif olaylar olarak iki kısma ayrılır. Hep aynı şekilde gerçekleşen olaylara tipik olaylar diyoruz. Örneğin, insanlar iki gözlü olarak doğması, Karadeniz bölgesi, İskandinav ülkeleri gibi bol yağış alan bölgelerde evlerin çatıların dik olması, Ekvator bölgesi insanının siyah derili olması, birer tipik olaydır. Tipik olaylar, hangi özelliklere sahip olduğu önceden bilinen olaylardır. Kollektif olaylar ise rassal olarak ortaya çıkan ve değişik sonuçlar alan olaylardır. Örneğin bir kentte yıllık kurumlar vergisi 20.000 tl üzerinde olan işletmeler, bir sınıftaki öğrencilerin boy ölçüleri birer kollektif olaydır. Bütün işletmeler 20.000 tl üzerinde vergi verseydi, bütün öğrenciler aynı boy, sözgelimi 1.70 cm, ölçüsüne sahip olsalardı o zaman bu olaylar da birer tipik olay haline dönüşürdü. Kollektif olayların sonucu, gözlemlenmeden, yani olay ortaya çıkmadan bilinemez.

Hemen sezinlenebileceği gibi, tipik olaylar istatistiksel araştırmanın konusu dışındadır. İstatistiksel araştırmanın konusu ancak kollektif olaylardır. Bütün öğrenciler aynı boyda ise, ailelerin gelir düzeyi hep aynı ise, her işletme aynı sayıda personel çalıştırıyor ise o zaman araştırılacak bir şey yok demektir. O zaman bir öğrencinin boyu, bir işletmenin personel sayısı, bir ailenin gelir düzeyi, bütün öğrencilerin boyu, bütün işletmelerin personel sayısı, bütün ailelerin gelir düzeyi hakkında yeterli bilgi verirdi. Başka bir izahla, tipik olaylar sabit, kollektif olaylar ise değişken olaylardır. Sabit olaylar istatistiğin konusu dışında kalırken, değişken olayların genel karakterini araştırmak ortaya koymak ise istatistiğin konusudur.

Veri ve Bilgi: Veri, belli bir konudaki gerçeklerin sembolik ifadesidir. Veri, bir gözlem veya faaliyet sonucu ortaya çıkarılabilir. Aynı konuda ve aynı zaman içinde birden fazla verilerin bizim için ifade ettiği toplam değere *Bilgi* diyoruz. Örneğin bir sınıftaki öğrencilerin boy ölçüleri elimizde olsun. Bunlar bizim verilerimizdir. Bu verilerden sözgelimi sınıfın boy ortalamasını hesaplıyorsak, bulacağımız değer bir bilgidir. Bir başka açıdan, veriye "*ham bilgi*", bilgiye ise "*işlenmiş veri*" diyebiliriz.

Birim: İstatistiksel bilgilerin derlendiği temel kaynağa birim diyoruz. Öğrencilerin boy ölçüleri araştırıldığında, birim her bir öğrencidir; ailelerin gelir düzeyi araştırıldığında, birimimiz her bir ailedir.

Yığın: Bir araştırmaya konu olan bütün birimlerin oluşturduğu topluluğa yığın diyoruz. Yığın yerine *Popülasyon veya Ana Kitle veya kısaca Kitle* denildiği de görülür. Yığının hangi birimlerden oluştuğu, yani yığının sınırları bilinmelidir. Ailelerin gelir düzeyleri araştırılırken belli bir kentteki ailelerin ya da Türkiye deki bütün ailelerin gelir düzeyi araştırılabilir. Bunlar araştırılan yığın için iki ayrı sınır tanımlar.

Örneklem: Araştırmalarda yığını oluşturan bütün birimlere ulaşmak, ya imkansız ya da çok masraflı olabilir. Bu durumda yığını iyi temsil edebilen bir alt küme seçilerek incelenir. Bu alt kümeden hareketle yığın için tahminde bulunuruz. İşte yığının bu alt kümesine *örneklem* adı verilir. Üniversitemizdeki öğrencilerin boy ortalaması ile ilgilendiğimizi düşünelim. Üniversitemizin bütün öğrencilerinin boylarını ölçmek yerine sabahleyin kampüse giren bir otobüste bulunan bütün öğrencileri bir örnek olarak alıp, bunların boy ölçülerinin ortalamasını yığının boy ortalaması yerine kullanabiliriz.

Değişken: Birimlerin araştırmaya konu olan her hangi bir özelliğine değişken (ya da vasıf) denir. Birimlerimiz her bir öğrenci olduğunda bunların boy ölçüleri, ağırlık ölçüleri birer değişkendir, öğrenciden öğrenciye değişir. Ailelerin gelir düzeyleri, sahip oldukları çocuk sayısı birer değişkendir. Değişkenleri, (1) sayısal (nicel, kantitatif) değişkenler (boy, ağırlık, gelir düzeyi gibi), (2) sayısal olmayan (nitel, kalitatif) değişkenler (cinsiyet, göz rengi, meslek gibi), olmak üzere ikiye ayırabiliriz. Ayrıca sayısal değişkenler sayı ekseninde belli bir aralıktaki her değeri alıp almamasına göre (1) sürekli değişken, örneğin boy, ağırlık ölçüleri, (2) kesikli değişken, örneğin ailelerin çocuk sayısı, evlerin oda sayısı gibi, olmak üzere ikiye ayırıyoruz. Başka bir açıdan bakılırsa, sürekli değişkenler ölçmeye tabi olurken, kesikli değişkenler sayma işlemine tabi olurlar.

Parametre: Doğruda doğruya yığına ait olan, yığın aritmetik ortalaması, yığın varyansı gibi, bilgilere parametre adı verilir. Bir yığın için bu gibi bilgiler tek bir değerden ibaret olduğundan parametreler *sabit* birer sayı olarak düşünülebilir. Parametrelerin hesaplanabilmesi için yığını oluşturan bütün birimlere ulaşmak gerekir. Fakat çoğu araştırmalarda yığının tamamına ulaşmak ya imkansız ya da çok pahalı olduğundan, parametreler çoğu kez bilinmeyen değerlerdir.

Örnek İstatistiği veya kısaca İstatistik: Örnekten elde edilen bilgilere örnek istatistiği veya kısaca istatistik adı verilir ki, bu yukarıda istatistik kelimesi için verilen anlamlardan üçüncüsüne karşı gelmektedir. İstatistikler yığın parametrelerinin örneklemdeki izdüşümleridir.

Aynı yığın içerisinde bir örnekten ötekine geçildiğinde istatistik değerleri de değişir.

VERİLERDE ÖLÇME DÜZEYİ

İstatistiğin temelinde olayları ölçme eyleminin yer aldığını belirtmiştik. Bu ölçme eyleminin hassasiyeti dört ayrı düzeyde gerçekleşebilir.

Bu düzeyler basitten karmaşığa doğru şöyledir: (1) **Sınıflama ölçme düzeyi**, (2) **Sıralama ölçme düzeyi**, (3) **Aralıklı ölçme düzeyi**, (4) **Oranlı ölçme düzeyi**. Olayları ya da nesneleri belirli kategorilere ayırmak suretiyle bunları sınıflama düzeyinde ölçmüş oluruz. Örneğin tüketim mallarını, dayanıklı ve dayanıksız tüketim malları olarak ayırmak, insanları medeni hallerine göre bekar, evli, dul gibi sınıflara ayırmak, birer sınıflama düzeyinde ölçmedir. Sınıflama ölçme düzeyi sadece birimleri birbirinden ayırt etmeye yarar. Birimler arasında büyüklük küçüklük gibi bir mukayese bilgisi vermez.

Olayları birbiri ile mukayese etmek için bunları belli bir özelliğe göre derecelendirmek gerekiyorsa sıralama düzeyinde ölçme ortaya çıkar. Örneğin bir işletmenin mamullerini kalite açısından birinci, ikinci, üçüncü kalite olarak ayırmak; öğrencilerin başarı düzeyini zayıf, orta iyi, pekiyi olarak belirlemek sıralayıcı ölçme düzeyidir. Sıralama düzeyindeki ölçme değerleri arasında küçük, büyük, eşit gibi mukayeseler yapılabilir. Sözelimi orta alan öğrencinin başarı düzeyi, iyi alan öğrencininkinden küçüktür. Ancak sıralama ölçme düzeyinde nesnelerin eşit olmaması yani büyük veya küçük olması durumunda bu farklılığın ne kadar olduğu sorusuna cevap verilemez, oranlama yapılamaz.

Aralıklı ölçme düzeyinde olayların veya nesnelerin belli bir özelliğe ne kadar sahip olduğu eşit aralıklı miktarlar olarak belirtilir. Aralıklı ölçmede başlangıç noktası mutlak değil, keyfidir. Aralıklı ölçmenin en tipik örneği ısı ve takvim ölçüleridir. Santigrat ile Fahrenheit, Miladi takvim ile Hicri takvim farklı başlangıç noktaları ile sırasıyla birer ısı ve zaman ölçüleridir. Aralıklı ölçme rakamları ile toplama veya çıkarma işlemleri anlamlı olurken çarpma ve bölme işlemleri anlamsızdır. Örneğin Malatya da ısı 18 santigrat Erzurum'da 9 santigrat ise aradaki ısı farkının 6 santigrat olduğunu miktar olarak söyleyebiliriz. Fakat bu ısı derecelerini örneğin birbirlerine oranlayarak bir sonuca gidilemez.

Aralıklı ölçme işleminde mutlak bir sıfır, yani başlangıç noktası varsa bu ölçme düzeyine oranlı ölçme düzeyi adı verilir. Örneğin uzunluk ölçülerinde, ağırlık ölçülerinde başlangıç noktası hep sıfırdır. Oranlı ölçme düzeyinde ölçülerin nisbi mukayesesi de mümkündür. Isı ölçülerinde örneğin Malatya Erzurum'dan yüzde şu kadar daha sıcaktır diyemeyiz ancak uzunluk ölçülerinde örneğin Ahmet'in boyu Mehmet'in boyunun yüzde sekseni kadardır diyebiliriz. Kısaca belirtmek gerekirse, başlangıç düzeyinin sabit olması ve oranlanabilir

olmaları oransal Ölçeklerin temel özelliğidir. Oranlı ölçme düzeyi rakamlarıyla bütün matematiksel işlemler yapılabilir.

TABLolar VE GRAFİKLER

İstatistiğin betimsel ve tümevarımsal olmak üzere iki kısma ayrıldığını belirtmiştik. Betimsel istatistiğin başlıca amacı, istatistiksel verilerin en iyi şekilde kullanıma sunulmasıdır. Veriler, tablo halinde veya grafik adı verilen şekillerle kullanıma sunulabilir.

Satır ve sütunlar halinde düzenli bir biçimde bir araya getirilmiş sayısal ya da sayısal olmayan bilgilerden meydana gelen topluluğa *Tablo* denir. Hemen her tabloda bir başlık (tablonun adı) ve alt başlıklar (satır veya sütun adları) bulunur. Bir tablonun esas kısmını meydana getiren satır veya sütun halindeki değerler dizisine *Seri* adı verilir.

SERİLER

Tabloda gösterilecek verilerin çokluğuna göre basit, sınıflanmış veya gruplanmış seri türlerinden biri seçilir. Basit seride veriler tek bir sütundan ibarettir. Verilerimiz az sayıda birimden oluşuyor ise bunları basit bir seri ile gösteririz.

ÖRNEK 1:

Bir öğrencinin çeşitli derslerden aldığı puanlar (X_i) aşağıda ki gibi basit bir seri oluşturur:

Dersler	Puan(X_i)
İstatistik	70
Muhasebe	65
İktisat	75
Matematik	80
İşletme	55
Hukuk	60
İngilizce	50

Verilerin basit seri olarak sunulması çok uzun bir liste oluşturuyorsa, seride tekrar eden değerleri sınıflayarak her sınıfın karşısına *frekans* adı verilen tekrar sayısı yazılmak suretiyle seri kısaltılabilir ki bu yeni seriye *sınıflanmış seri* adı verilir. Şu halde sınıflanmış serilerde, biri değişken değerleri sütunu ve diğeri frekans sütunu olmak üzere iki sütun mevcuttur.

ÖRNEK 2:

Bir sınıfta 20 öğrencinin İstatistik dersi ara sınav puanları şöyledir: 50, 55, 70, 45, 50, 30, 70, 40, 55, 40, 35, 65, 40, 35, 55, 60, 40, 75, 40, 55. Bu 20 öğrenciye ait puanları basit seri olarak göstermek uzun bir liste oluşturur. Bu verileri sınıflanmış seri olarak daha derli toplu bir biçimde sunabiliriz. Tekrar eden değerler sınıflandırılırsa:

Puan (X_j)	Öğrenci Sayısı (f_j)
30	1
35	2
40	5
45	1
50	2
55	4
60	1
65	1
70	2
75	1
Toplam	20

Birim sayısı daha da arttığında yukardaki örnekte alınan farklı puanların sayısı artma eğiliminde olacak ve buna bağlı olarak tablodaki sınıf sayısı da artacaktır. Bu durumda seriyi gruplandırmamız ve her gruba düşen birim sayısını göstermemiz gerekir. Bu şekilde *gruplanmış seri elde* edilir.

ÖRNEK 3: Bir sınıfta İstatistik ara sınavı puan dağılımı şöyledir:

Puanlar(X_j)	Öğ. Sayısı (f_j)
$0 < X \leq 20$	6
$20 < X \leq 40$	7
$40 < X \leq 60$	15
$60 < X \leq 80$	13
$80 < X \leq 100$	9
Toplam	50

Tablodaki X , puan değişkenini simgelemektedir. Görüldüğü gibi gruplanmış seri de sınıflanmış seri gibi, değişken ve frekans sütunlarından oluşmaktadır. Sınıflanmış seride

değişken sütunu tek değerlerden oluşurken gruplanmış seride değişken sütunu değer gruplarından meydana gelmektedir. Gruplanmış bir seri oluştururken dikkat edilecek hususlar şunlardır:

- Değişkenin alabileceği her değere gruplamada yer verilmelidir.
- Değişkenin her değeri sadece ve sadece bir gruba dahil olmalıdır.

Örneğin, yukarda 80 puanı sadece dördüncü gruba dahildir.

-Gruplandırmada her grubun uzunluğu aynı olmalıdır.

Örneğimizde her grubun uzunluğu 20 birimdir. Bu şekilde verilerin tablo halinde sunulmasında eldeki verilerin çokluğuna göre bu üç değişik gösterimden en uygun (en derli toplu) olanını seçmek durumundayız.

Gruplanmış seri özellikle sürekli değişkenler için uygun olurken, sınıflanmış seri kesikli değişkenler için daha uygundur. Bazı gruplanmış serilerde ilk grubun alt sınırı ile son grubun üst sınırı açık yani sınırsız olabilir. Bu tür serilere *açık uçlu seriler* adı verilir. Bu seriler tablo olarak sunulabilir de, istatistiksel işlemlere girişebilmek için bu açık uçlara belli bir değer verilmesi zorunludur. Sınıflanmış ve gruplanmış serilere genel olarak *frekans serileri* de denilmektedir.

KÜMÜLATİF SERİLER

Basit bir seride seri değerlerini aşağı veya yukarı doğru biriktirmek suretiyle *kümülatif seri* elde edilir. Frekans serilerinde ise, hem seri değerlerini hem de frekansları aşağı veya yukarı doğru biriktirmek suretiyle *kümülatif frekans serisi* elde edilir.

ÖRNEK 4:

Aylar itibariyle bir ülkenin ihracatı basit bir seri ve kümülatif seri olarak gösterilmiştir:

Aylar	İhracat (Milyon-\$)	Kümülatif Seri
Ocak	43	43
Şubat	47	90
Mart	51	141
Nisan	50	191
Mayıs	55	246
Haziran	58	304
Temmuz	60	364
Ağustos	56	420
Eylül	62	482
Ekim	65	547
Kasım	63	610
Aralık	70.	680

Kümülatif seri değerleri, kullanıcıya o noktaya kadar olan toplam değeri vermektedir. Örneğin yıl içerisinde Mayıs ayına kadar yapılan toplam ihracat 246 milyon dolar olduğu görülüyor.

NİSBİ SERİ

Yukardaki serilerde verileri mutlak rakamlarla sunmuş olduk. Ancak verileri kullananlar bazen rakamların mutlak büyüklüğü ile değil de bunların nisbi büyüklüğü ile ilgilenebilirler. Bunun için normal serileri nisbi serilere dönüştürmek gerekir. Basit bir seride seri değerleri seri toplamına oranlanmak suretiyle *nisbi (oransal) seri* elde edilir. Frekans serilerinde ise her bir frekansın toplam frekansa oranlanması ile elde edilen yeni seriye *nisbi frekans serisi* adı verilir.

Bir frekans serisindeki sınıf veya grup sayısı m tane ise Toplam frekans $N = \sum_{j=1}^m f_j$ ile

gösterilebilir. Buna göre j inci sınıf veya grup için, Nisbi frekans $p_j = \frac{f_j}{N}$, $j = 1, \dots, m$ ile formüle edilebilir.

ÖRNEK 5:

(4) no'lu örnekteki ihracat verileri basit nisbi serisini oluşturalım. Her aylık değer toplam değere bölünerek oranlar, bulunur:

Ocak	$43:680 = 0.063$
Şubat	$47:680 = 0.069$
Mart	$51:680 = 0.075$
Nisan	$50:680 = 0.074$
Mayıs	$55:680 = 0.081$
Haziran	$58:680 = 0.085$
Temmuz	$60:680 = 0.088$
Ağustos	$56:680 = 0.082$
Eylül	$62:680 = 0.091$
Ekim,	$65:680 = 0.096$
Kasım	$63:680 = 0.093$
Aralık	$67:680 = 0.098$

Bu hesaplamalar 100 üzerinden okunabilir. Buna göre örneğin, Mayıs ayında yıllık toplam ihracatın % 8.1 'i Haziran ayında yıllık toplam ihracatın % 8.5'i yapılmıştır. Nisbi seri aşağıdaki şekilde, düzenlenebilir.

Aylar	Oranlar
Ocak	0.063
Şubat	0.069
Mart	0.075
Nisan	0.074
Mayıs	0.081
Haziran	0.085
Temmuz	0.088
Ağustos	0.082
Eylül	0.091
Ekim	0.096
Kasım	0.093
Aralık	0.098
Toplam	100.0

NİSBİ KÜMÜLATİF SERİ

Basit kümülatif seri değerleri seri toplamına bölünmek suretiyle *nisbi kümülatif seri* bulunur, m sınıf ya da gruplu bir kümülatif frekans serisinde

F. her bir frekans değerinin toplam frekansa oranlanması, $P_j = \frac{F_j}{N}$, $j = 1, \dots, m$ ile

bulunan yeni serilere de *nisbi kümülatif frekans serileri* adı verilir. Bu seri türleri için birer örnek verelim.

ÖRNEK 6:

Dört nolu örnekteki veriler için basit kümülatif oranlar:

Ocak	43:680 = 0.063
Şubat	90:680 = 0.132
Mart	141:680 = 0.207
Nisan	191:680 = 0.281
Mayıs	246:680 = 0.362
Haziran	304:680 = 0.447
Temmuz	364:680 = 0.535
Ağustos	420:680 = 0.618
Eylül	482:680 = 0.709
Ekim	547:680 = 0.804
Kasım	610:680 = 0.897
Aralık	680:680 = 1.000

Buna göre ihracat rakamlarının nisbi kümülatif serisi tablo olarak:

Aylar	Nisbi Küm. Frekans
Ocak	0.063
Şubat	0.132
Mart	0.207
Nisan	0.281
Mayıs	0.362
Haziran	0.447
Temmuz	0.535
Ağustos	0.618
Eylül	0.709
Ekim	0.804
Kasım	0.897
Aralık	1.000

Basit nisbi kümülatif seri herhangi bir zamana kadar olan toplam yüzdelik miktarı vermektedir. Bu tablodan örneğin Eylül ayı itibariyle, yıllık ihracatın % 70.9 nun yapıldığı anlaşılmaktadır.

Daha önce ihracat verileri için çeşitli seriler bulunmuştu. Bu serilere ilişkin diyagramlar çizilebilir. Aylar itibariyle kümülatif ve nisbi kümülatif serileri tekrar aşağıdaki tabloda gösterilmiştir:

Aylar	İhracat	Kümülatif	Nisbi Kümülatif
1	43	43	0.063
2	47	90	0.132
3	51	141	0.207
4	50	191	0.281
5	55	246	0.362
6	58	304	0.447
7	60	364	0.535
8	56	420	0.618
9	63	482	0.709
10	-65	547	0.804
11	63	610	0.897
12	67	680	1.000