

1. 1. İstatistik Nedir?

Tanım 1: İstatistik, belirli bir amaç için veri toplama, tablo ve grafiklerle özetleme, sonuçları yorumlama, sonuçların güven düzeylerini açıklama, örneklemelerden elde edilen (hesaplanan) sonuçları kitle için genelleme (çıkarsama), karakteristikler arasındaki ilişkileri araştırma, çeşitli konularda geleceğe ilişkin projeksiyon yapma, deney tasarlama ve düzenleme v.b. kapsayan bir yöntemler topluluğudur.²

Tanım 2: İstatistik verileri işler ve özetler, araştırmacılara yol gösterir. Değişkenler arasındaki ilişkileri incelememize, tahminler ve öngörüler yapmamıza, doğru karar vermemize yarar.³

Tanım 3: Belirli amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimi olarak tanımlanabilir.⁴

Tanım 4: İstatistik; gözlem ve ölçme sonucunda elde edilen sayısal (rakamsal) verilerin tablolar veya grafikler halinde sunulması ve bunların karar alma sürecinde nasıl kullanılacağına ilişkin yöntemler sunar. İstatistiğe konu olacak olaylar gözlemlenebilir ve ölçülebilir olmalıdır. İstatistikler belirli ölçümler yapılarak elde edilmişlerdir yani istatistikler rastgele bulunmuş sayılar değildirler. Belirli bir kitleyi oluşturan birimlerin gözlemlenmesiyle elde edilen, ölçülebilen veya sayılabilen nicel bilgileri ifade etmektedirler. Bu nedenle; logaritma cetveli, telefon rehberi, milli piyango sonuçları birer istatistik değildirler.

Tanım 5: İstatistik; olasılığa dayalı yani sonucun baştan kesin olarak bilinmediği rastlantısal (tesadüfi) durumlar üzerinde çalışırken, belirsizlik altında karar vermeyi sağlar.

Tanım 6: Verilerin derlenmesi, işlenmesi ve yorumlanmasına yönelik yöntemlerin tümü. (İşleme ve yorum verileri temsil edebileceği düşünülen kuramsal dağılım ve modellere geçişi de içermek üzere matematiksel istatistik kapsamına girer)⁵. İstatistiksel araştırmalara birkaç örnek verecek olursak, örneğin, bir araştırmacı, hamile iken diyetisyen kontrolünde iyi ve kaliteli beslenmenin doğacak bebeğin genel sağlık ölçümlerinde, düzensiz ve bilinçsiz beslenen bir annenin bebeğine kıyasla olumlu bir etkisi var mıdır? Fazla kilolu olmanın yaşam süresi üzerinde gerçekten olumsuz etkileri var mıdır? Sigara kanser ilişkisinin boyutu nelerdir? Kanser olmayı tetikleyen diğer faktörler neler olabilir ve risk sıralaması nasıldır?

² (Özer Serper, *Uygulamalı İstatistik 1 -2*)

³ (Necmi Gürsakal, *Betimsel İstatistik*, syf. 6)

⁴ (Murat Açıköğretim Yayınları, *İstatistik*)

⁵ (Büyük Larousse, Cilt: 10, Syf: 5872)

1.2. İstatistiğin Temel Konuları

İstatistik konuları bakımından ikiye ayrılır; Betimsel (tasviri) istatistik ve Çıkarımsal istatistik.

1.2.1. Betimsel İstatistik

Verilerin toplanması, yorumlanıp özetlenmesi gibi konularla ilgilenir. Tanımsal istatistik, verilerin bir takım sayısal ve(ya) grafiksel yöntemlerin kullanımı ile bilgiye dönüştürülmesini ve karar verme sürecinde kullanılmasını sağlar.⁶

İstatistik, geçmiş ve içinde bulunulan durumu tanımlayarak bir veri kümesine ilişkin özet değerler ve grafikler ortaya koyduğunda betimsel istatistik (descriptive statistics) adını alır. Özetleme ve verilerin yoğunlaştırılması diğer bir deyişle verilerin hacimleri azaltılarak kullanım değerlerinin arttırılması betimsel istatistiğin konusuna girer. Tablolar ve grafikler yardımı ile verilerin özetlenmesi ve çok sayıda sayıdan oluşan bir veri grubunun “ortalama” gibi tek bir sayıya indirgenmesi yine bu alan için geçerlidir. Kısaca betimsel istatistik bir veri kümesinde bulunan bilgiyi sayısal ve grafiksel yöntemleri kullanarak özetler ve sunar.⁷

Betimsel amaçlı istatistik kitledeki tüm birimlerden ilgili değişken ya da değişkenler için veri toplandığında bunları kullanarak kitleyi özetlemeyi (betimlemeyi) amaçlar. Bu ise frekans dağılımı oluşturularak, grafikler çizilerek ya da parametreler (kitle ortalaması ve varyansı) hesaplanarak yapılır.⁸

1.2.2. Çıkarımsal İstatistik

Tümevarımsal amaçlı istatistik kitle(ler)den rastgele seçilen örneklem(ler)den toplanan verileri kullanarak kitle(ler) nin parametrelerini tahmin etmeyi veya parametrelerle ilgili olan savların doğru olup olmadığının araştırılmasını amaçlar. Yanlı örneklem gözlemlendiğinde belirsizlikler içeren kitle hakkındaki önermelere geçmek için geliştirilen süreçle istatistiksel sonuç çıkarır.⁹

Çıkarımsal istatistik, araştırma sürecinin betimsel istatistiğin bıraktığı yerden devam ettirilmesi suretiyle, çalışmanın amacına uygun bir şekilde daha ileri tekniklerle (olasılık başta olmak üzere) karar verme ve kitleye dair çıkarsamalar yapma sürecidir. Bu çıkarsamalar örnekleme dayalı olarak yapıldıkları için belirli bir hata düzeyinde ifade edilirler. (Kitle N adet birimden oluşan teorik topluluktur, aslında tüm birimlere ulaşmak olasıdır ancak zaman ve

⁶ Onur Özsoy, İktisatçılar ve İşletmeciler İçin İstatistik, Excel Uygulamalı, 2010, Ankara, Syf: 4.

⁷ Necmi Gürsakal, Betimsel İstatistik, Syf:20.

⁸ Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.

⁹ Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.

maliyet kısıdından dolayı genellikle istatistiksel çalışmalar kitleyi iyi bir biçimde temsil eden daha küçük alt gruplar yani örneklemeler üzerinden yürütülür.

Örneklemin büyüklüğü n ile gösterilir. O halde N büyüklüğündeki bir kitleden n büyüklüğünde tane mümkün örneklem çekilebilir($C_n^N = \frac{N!}{(N-n)!n!}$). Bu örneklemeler büyüklük yani hacim olarak birbirinin aynı fakat içerdikleri birimler bakımından birbirinden farklıdır, kombinasyonun doğası gereği seçilen birim yerine konmamaktadır, bu nedenle mümkün örneklem sayısı hesaplanırken permütasyon yerine kombinasyon kullanılmaktadır.

Örnekleme dayanan çıkarımlar mutlak doğru olamayacağından “**olasılık**” kelimesi kullanılarak ifade edilir. Çıkarımsal istatistiğin genel olarak kullandığı yöntem “tümevarım” dır. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy lar boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.¹⁰

****Her şey sabit, belirli ve kontrol altında olsaydı, yani değişkenlik olmasaydı istatistiğe gerek kalmazdı!!!*

Yukarıdaki cümle istatistiğin özünü ifade etmektedir. Her şey bir ihtiyacı gidermek için ortaya çıkıp kabul gördüğüne göre, istatistik de, gerçek hayatı olduğu gibi ele alabilmeyi sağlamıştır. Olasılık teorisinin katkılarıyla belirsizliği ölçebilmeyi, geleceğe dair tahminler ve çıkarımlar yapabilmeyi olanaklı kılmıştır. Yaşamın belirsizlikleri insanoğlunu her zaman 2 kere 2'nin dört etmeyeceği gerçeğiyle karşı karşıya bırakmıştır. Özellikle sosyal bilimlerde olayları salt pozitif bilimlerle açıklamak yetersiz kalmaktadır. İstatistik böyle durumlarda ölçümlemeyi sağlayabilen yöntem ve teknikler bütünü olarak öne çıkmaktadır.

1.3. İstatistiğin Tarihçesi

“İstatistik kelimesi Almanya’da Achenwall tarafından 1748’de kullanılmaya başlanmışsa da, istatistik çalışmaları çok eski çağlara kadar uzanır. Tacitus, imparatorluk zenginlikleri üzerinde Augustus’un geniş bir anket yaptığını yazar. Böylece askerler, gemiler her çeşit gelir kaynakları sayılmış ve kamu gelirleri tespit edilmiştir.

Bütün Ortaçağ boyunca ve 17. Yy’a kadar istatistik sadece tasviri olarak kalmıştır. Bu alanda iki okul ortaya çıktı: Göttingen Üniversitesi profesörü ve Almanlar tarafından istatistiğin babası olarak bilinen Achenwall’in temsil ettiği tasviri okul ve bazı sosyal olayların yaklaşık düzenliliğinden öngörüler ve kanunlar çıkarmayı deneyen siyasi matematikçiler okulu. 18. Yy’da Fransız Desparcieux ve İsviçreli Wargentın toplum olaylarını öngörmenin pratik önemini göseren ve çok gelişmekte olan sigorta sanayinin hareket noktasını teşkil eden ilk ölüm tablolarını (mortalite tablosu) düzenlemişlerdi.

Fransa’da ilk resmi istatistik bürosu 17. Yy’ın sonlarında kurulmuş ve içişleri bakanı Chaptal 1801’de genel nüfus sayımını gerçekleştirmişti. Türkiye’de ilk resmi istatistik bürosu 1933’te kuruldu.

¹⁰ Özer Serper, Uygulamalı İstatistik 1- 2.

Jacques Bernoulli ve özellikle Laplace, tasviri sayısal bilgilere matematik bilgileri ilave ederek sonuçların değerlendirilmesinde olasılık hesap imkânlarının kullanılmasını araştırdılar. Birincisi ünlü “büyük sayılar kanunu” nu ortaya attı, ikincisi ise “analitik olasılık teorisi” adlı eserinde karmaşık sebepli doğal olayların incelenmesinde olasılık hesaplarından sağlanacak faydaları açıkça belirtti. Quotelet, çevresinde bütün diğer insanların yer alacağı hayali bir “ortalama insan” bulmak maksadıyla bu yöntemin uygulanışını insanların fiziki, ahlaki ve düşünsel özelliklerine uyguladı.” ¹¹

“Quotelet bütün sosyal olayları birey etrafında toplamıştır. Daha ileri giderek istatistiği bütün sosyal alanlara hatta ahlaki olaylara da yaymıştır. İlk “sosyal kriminolog” olarak adlandırılmaktadır. 1836 yılında Quotelet, “Toplum suçu hazırlar, Suçlu ise ancak bir araçtır” demiştir. Bugün Quotelet modern istatistiğin kurucusu olarak kabul edilmekle birlikte, ileri sürdüğü sistem ve vardığı sonuçlar reddedilmektedir. Quotelet ’ten sonraki gelişmeler istatistik yöntembiliminde matematiğin daha yaygın bir şekilde uygulanmasını sağlamış ve matematiksel istatistik disiplininin meydana çıkmasına yol açmıştır.

İstatistik; Türklere de yabancı bir bilim dalı değildir. Selçuklular döneminde İran’da, İlhanlılar döneminde Hint’te nüfus sayımları düzenlenmişti. Öte yandan Osmanlı İmparatorluğu’nun yönetim sistemi nüfus ve arazi hakkında düzenli biçimde bilgi toplamasını gerektiriyordu. Bu nedenle imparatorluk zamanında 30-40 yıl gibi aralıklarla nüfus sayımları ve arazi yazımları yapılmaktaydı. Bu istatistiksel işlemler sırf vergi toplanması amacıyla yapıldığı için sadece vergi mükellefleri göz önünde bulunduruluyordu. İmparatorluğun gerileme devrinde bu sayımların arkası kesilmiş, modern tekniklere dayanan istatistikler de 20. Yy’e kadar ülkeye girememiştir. Gerçi bu gibi istatistiklerin düzenlenmesi yolunda 19. Yy’de bazı çalışmalar yapılmıştır. Örneğin orduyu modernleştirmek amacıyla askere alınabilecek erkeklerin sayısı öğrenilmek istenmiş ve böylece ilk sayım 1831’de gerçekleştirilmiştir. Ancak bu sayım yalnız erkekleri dikkate aldığı gibi ülkenin bütünü de kapsamamıştır. Nitekim sayım yapılan yerler Rumeli ve Anadolu sancakları ve kasabalarıdır. Dolayısıyla bu sayım bugünkü anlamıyla gerçek bir nüfus sayımı olmaktan uzak kalmıştır. 1844’te yapılan ikinci nüfus sayımının amacı ise kimlik belgesi verilecek vatandaşların belirlenmesi idi. Bu nedenle kadınlar da erkeklerle birlikte sayılmıştır. Ne var ki askerlik korkusuyla birçok erkek sayım dışında kalmıştır.

19. Yy’de ülke nüfusu hakkında bilgi verecek sonucu sayım 1874’te başlamış ve araya giren Rus savaşı nedeniyle 1884’e kadar sürmüştür. Nüfus kütüklerini düzenleme ve halka nüfus tezkeresi verme amaçlarıyla başlatılan bu sayımın uygulanmasında da bazı aksaklıklar görülmüştür. Birçok teknik hata içeren ve kısmi nitelikteki bu sayımlar yanında ölüm ve doğumlar gibi nüfus hareketlerini suçları ithalat ve ihracatı belirlemek amacıyla çalışmalar yapılmıştır. Ne var ki bu çabalar da sonuçsuz kalmıştır.” ¹²

¹¹ *Meydan Larousse, Cilt: 6, Syf: 511-512)*

¹² *(Özer Serper, Uygulamalı İstatistik 1 – 2)*

1.3.1. Tarihten Günümüze Öne Çıkan İstatistiksel Çalışmalar

- Antikçağ'da Çin'de, Mısır'da, Yunanistan'da nüfus ve bu nüfusun maddi yaşam koşullarına ilişkin sayısal verilere sahip olma gereksinimi izlerine rastlanmıştır.
- Roma'da düzenli nüfus sayımlarına rastlanır. Roma'da önce 5 yıl, sonra da 10 yıl ve 15 yıl gibi düzenli aralıklarla yapılan ve "census" adı verilen bu sayımlar 600 yıl kadar sürmüştür. Her Romalı kendi ve babasının adını, servetini, arazisini, kölelerinin sayısını bildirmek zorundaydı.¹³
- 1662'de Graunt erkek çocuklarının doğumlarının kız çocuklarının doğumlarına oranı gibi değişmezlikleri ortaya çıkardı.
- 1742'de Edmond Halley çağdaş aktüer (hayat sigortacılığı) çalışmalarının temelini oluşturan bir hayat tablosu yaptı.
- 1767'de Süssmilch erkek doğum oranı ve 20 yaşına dek erkek çocuk oranındaki gelişme üzerine önemli çalışmalar yayınladı.
- Pierre Simon De Laplace (1749-1827) "Çözümsel Olasılık Kuramı" (1812) adlı eserinde, nedenlerinin tümünü bilip tek tek çözümlenemeyecek kadar karmaşık olan doğal olayların incelenmesinde bu kuramdan sağlanacak yararları ortaya koydu. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.¹⁴
- Adolphe Quetelet (1796-1874) yöntemin uygulama alanını, canlı varlıkların antropometrik, psikolojik ve toplumsal açıdan incelenmelerini sağlamıştır. Onun girişimiyle 1885'te Londra'da kurulan Uluslararası İstatistik Enstitüsü'nün öncüsü olan ilk uluslararası istatistik kongresi 1853'te Brüksel'de toplanmıştır.
- Karl Pearson (1857-1936) Biyoistatistik'i buldu. Bu yıllarda istatistik ile iktisadın arasındaki ilişkilerin fark edilmesiyle ekonometri doğmuş oldu.¹⁵

¹³ Özer Serper, Uygulamalı İstatistik 1-2.

¹⁴ A.g.e.

¹⁵ (Büyük Larousse, Cilt: 10, Syf: 5873)

1.4. Türkiye İstatistik Kurumu Hakkında Genel Bilgiler

“Cumhuriyet döneminde istatistik işlerine büyük ölçüde önem verilmeye başlanmıştır. Nitekim 1926’da Merkezi İstatistik Dairesi kurulmuştur. Kurumun adı adeta yazboz tahtası yapılarak 1930’da İstatistik Umum Müdürlüğü, 1945’te İstatistik Genel Müdürlüğü, 1952’de İstatistik Umum Müdürlüğü, 1960’ta İstatistik Genel müdürlüğü şekline çevrilmiş ve 1962’de Devlet İstatistik Enstitüsü adında karar kılınmıştır. 2005 yılında ise Türkiye İstatistik Kurumu (TÜİK) kurulmuş olup, kurum istatistik konseyi ve Türkiye İstatistik Kurumu başkanlıklarından oluşmuştur. TÜİK Başkanlığı 5429 sayılı Türkiye İstatistik Kanunu’nun uygulanması sağlamak ve kanunla kendisine verilen görevleri yerine getirmek üzere kurulmuş olup, merkez ve taşra örgütlerine sahiptir. Merkez örgütü ana hizmet danışma ve yardımcı hizmet birimlerinden oluşmaktadır. Taşra örgütü ise bölgesel düzeydeki tüm istatistiksel faaliyetleri yürütmek ve yerel birimlerle koordinasyonu sağlamakla görevli bölge müdürlüklerini kapsamaktadır.”¹⁶

“Enstitü gerekirse Ankara dışında geçici veya sürekli bürolar kurma ve mahalli muhabirler kullanma yetkisine sahiptir. Görevleri: her çeşit çalışmalara ilişkin istatistikleri düzenler, istatistikle ilgili olarak kamu kuruluşları arasında koordinasyon saplar, çözüm ve incelemeler yapar, düzenlenen istatistikleri yayınlar, istatistik yaylığı çıkarır, sayımlar ve anketler hazırlar ve sonuçlarını yayınlar, kamuoyu çalışmaları yapan bilimsel kurumların çalışmalarında yardım eder, yurt içi ve yurt dışı istatistik eğitim işleri yapar.”¹⁷

Kurumun adı 2005 yılında Türkiye İstatistik Kurumu (TÜİK) olarak değiştirilmiştir ve halen bu isimle anılmaktadır.

1.5. Günümüzde İstatistik

ASA (American Statistical Association) 2013 yılını dünya İstatistik yılı ilan etmiştir. North Carolina Üniversitesi’nden Marie Davidian’ın 19 Aralık 2013’te JASA (Journal of American Statistical Association) dergisinde yayınlanan “The International Year of Statistics: A Celebration and A Call to Action” adlı makalesinde istatistiğin dünyada yükselen bir değer olmasından bahsetmektedir.

Son yıllarda gazetelerde istatistikle ilgili başlıklar atılmıştır, birkaç örnek vermek gerekirse;

-“For today’s graduate, just one word: Statistics”, The New York Times, 2009.

-“The allure of the statistics field grows”, The Boston Globe, 2012

-“What are the odds that stats would be this popular?”, The New York Times, 2012.

-“The numbers guy”, The Wall Street Journal, 2013.

¹⁶ A.g.e.

¹⁷ Meydan Larousse, Cilt: 6, Syf: 512.

Öne çıkan yazılar bunlar olmakla beraber, istatistik biliminin yükselişı çok sayıda makaleye konu olmuştur.

İstatistiğin doğru kullanımı ve gücü hakkında son yıllarda en büyük katkısı 2012 Amerika devlet başkanlığı seçimlerinden önce yaptığı tutarlı tahminlerle Nate Silver yapmıştır. “The Signal and the Noise: Why So Many Predictions Fail But Some Don’t?” adlı kitabı bir anda en çok satanlar arasına girerek, tüm siyasilerin ve halkın ilgi odağı haline gelmiştir.

Ülkemizde çok uzak değil, bundan 10-15 yıl önce istatistik denilince akla, futboldaki özet sonuçlar veya sokakta yapılan anketler gelmekteydi. Oysa günümüzde tıpkı dünyada olduğu gibi istatistik hayatın her alanına girmiş ve etkin bir biçimde kullanılmaya başlanmıştır.

Bölüm Soruları

1. İstatistik nedir?
2. Betimsel istatistik nedir?
3. Çıkarımsal istatistik nedir?
4. İstatistiğin kullanım amaçları nelerdir?
5. İstatistiğin doğuşu ve tarihsel gelişimi, dünyadaki son durumu ile ilgili gelişmeler nelerdir?

2. İSTATİSTİKSEL ÇIKARSAMA SÜRECİ VE İSTATİSTİKTE TEMEL KAVRAMLAR

Bu Bölümde Neler Öğreneceğiz?

- İstatistiksel çıkarsama süreci
- İstatistikte temel kavramlar
- Değişken türleri
- İstatistikte hata türleri

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Çıkarımsal istatistik süreci, temel kavramlar ve hata türleri	İstatistiksel çıkarsamanın yapılışı ve temel kavramların öğrenilmesi	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

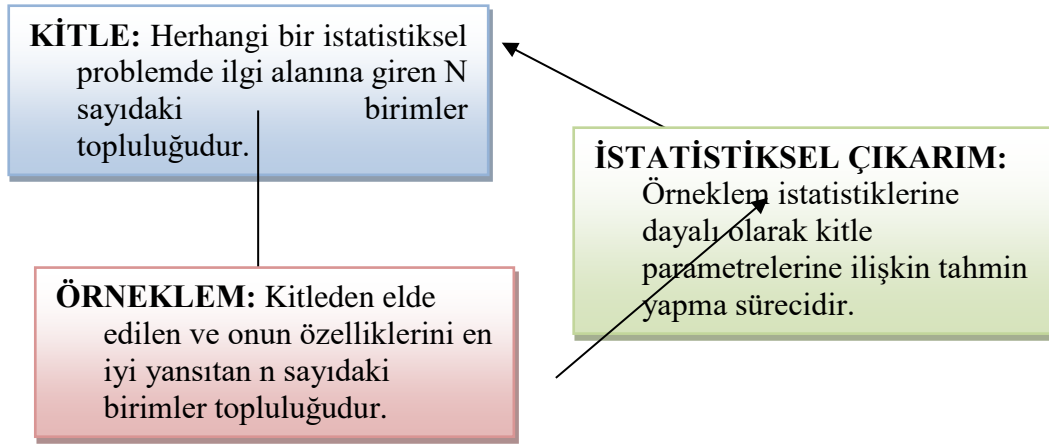
Anahtar Kelimeler

- İstatistiksel çıkarsama
- Kitle
- Parametre
- Örneklem
- İstatistik
- Birim
- Değişken
- Nominal veri
- Ordinal veri
- Rastsal hata
- Sistematik hata

2.1. İstatistiksel Çıkarsama Süreci

Örneklemden elde edilen bilgilere dayalı olarak kitleyle ilgili tahmin yapma ve karar verme sürecidir. İstatistiksel çıkarım istatistiğin temel amaçlarından birisidir. Yöneticiler veya karar verme mekanizmaları istatistiksel çıkarım yöntemleri ile aldıkları kararları ve geleceğe dönük olarak yaptıkları tahminleri test ederler. İstatistiksel çıkarım yöntemlerini kullanan yöneticiler firmanın gelecekte üretim düzeyinin ne olacağını, önceki dönemlere ait verileri inceleyerek ve mevcut piyasa koşullarını göz önünde bulundurarak tahmin edebilirler. Yapılan tahminlere dayalı olarak firmanın yönetim, pazarlama, tedarik ve reklam gibi fonksiyonları ve diğer stratejilerine ilişkin kararlar alınabilir. İstatistiksel çıkarım süreci aşağıdaki gibidir:

Şekil.2.1. İstatistiksel Çıkarım Süreci



2.2. İstatistikte Temel Kavramlar

Kitle:

Populasyon, yığın ve anakitle (anakütle) şeklinde de adlandırılan kitle, herhangi bir istatistiksel problemde ilgi alanına giren konu ile ilgili birimlerin tamamının oluşturduğu topluluktur. Büyüklüğü “N” ile gösterilir yani kitle N tane birimden oluşmaktadır.

Bir diğer tanım; kitle, hakkında bilgi edinilmek istenen ve biçimsel homojenliğe sahip (belirli bir tanıma uyan) kolektif olay niteliğindeki birimlerin oluşturduğu topluluğa denir.

Parametre:

Kitle ile ilgili ölçümlere parametre denir. Örneğin kitleye ilişkin varyans (σ^2), standart sapma (σ) ve ortalama (μ) birer parametredir. Bir kitle için bu gibi bilgiler tek bir değerden ibaret olduğundan, parametreler sabit bir sayı gibi düşünülebilirler. Parametrelerin hesaplanabilmesi için kitledeki her bir birimin hesaba alınması yani tamsayım yapılması

gerekir. Ancak bu durum çoğu kez tercih edilmez, zira hem maddi olanaklar bakımından hem de zaman sorunu göz önüne alındığında, parametreler çoğu kez bilinmeyen, ancak tahmin edilen değerlerdir.

Örneklem:

Kitleden elde edilen ve kitlenin özelliklerini en iyi yansıtan “n” adet birimden oluşan birimler topluluğudur. Kitlenin olası alt kümelerinden her birine örneklem denir. Araştırmalarda kitleyi oluşturan bütün birimlere ulaşmak bazen imkânsız çoğu zaman ise masraflı olabilir. Bu kitleyi iyi temsil edebilen bir alt küme seçilerek incelenir. Bu alt kümeden hareketle kitle için tahminlerde bulunuruz. İşte kitlenin bu alt kümesine “örneklem” adı verilir. Ekonometri bölümü 1. Sınıf öğrencilerinin ağırlıklarını ölçmek istediğimizde, tüm öğrencilerin kilosunu öğrenmek yerine rastsal olarak bir örneklem çekerek de yaklaşık bir değere ulaşabiliriz. Şüphesiz, parametrenin yansız bir tahminin yapmak yani gerçek değerine çok çok yakın bir istatistik elde etmek için seçeceğimiz örneklem kitleyi çok iyi temsil etmesi ve yeterli büyüklükte olması gerekmektedir. Söz gelimi, 500 öğrencinin ortalama ağırlığını tahmin etmekten bahsediyorsak, 10 öğrencinin ağırlığını ölçerek böyle bir tahmin yapmak uygun olmayacaktır. İlerleyen derslerde ideal örneklem büyüklüğünün nelere bağlı olarak belirlendiğini ve kaç olması gerektiğini öğrenmiş olacaksınız.

İstatistik (örneklem istatistiği):

Örnekleme ilgili tanımsal ölçüme denir. Örneğin örneklem varyansı, standart sapması ve örneklem ortalaması birer istatistiktir. Aynı kitle içinde bir örneklemden diğerine geçildiğinde istatistiklerin aldıkları değerler de değişmektedir. N büyüklüğündeki bir kitleden n hacimli mümkün tüm örnekler çekilse ki bu örneklemelerin sayısı N’in n’li kombinasyonu kadar sayıdadır, her biri için birer ortalama, varyans, vb. hesaplanabilir. Parametreler tek bir değer alırlarken, istatistikler örneklemden örnekleme değişen farklı değerler alabilmektedirler. Bu durumu kısaca bir örnekle açıklayalım. N=5 büyüklüğünde bir kitle olsun ve birimler aşağıdaki gibi olsun:

1 3 5 7 9

Kitle ortalaması, bütün bu birimlerin toplamının N’e yani 5’e bölünmesiyle elde edilir ve bu da 5’tir. Şimdi bu kitle ortalamasını örneklemeler üzerinden tahmin etmeye çalışalım. n=3 büyüklüğünde mümkün tüm örneklemelerin sayısı 5’in 3’lü kombinasyonu kadardır yani 10 adettir.

<u>Örneklemler</u>	<u>Örneklem Ortalaması</u>
1 3 5	3
1 3 7	3,66
1 3 9	4,33
1 5 7	4,33
1 5 9	5
1 7 9	5,66
3 5 7	5
3 5 9	5,66
3 7 9	6,33
5 7 9	7

Örneklem ortalamalarının ortalaması yani mümkün tüm örneklemlerden hesaplanan örneklem ortalamalarının ortalamasını hesaplırsak 4,997 elde ederiz ki bu da kitle ortalamasına çok yakın bir değerdir. Bu örneği vermekteki amacımız parametrenin örneklemler yoluyla nasıl tahmin edildiğine dair çok temel ve basitçe bir açıklama ihtiyacından ileri gelmektedir. Gerçekte böylesi az sayıda birimden oluşan bir kitle söz konusu olduğunda zaten örnekleme yapılmaksızın her birime kolayca ulaşılabileceği için doğrudan doğruya parametreler hesaplanabilir. Bu örnekten anlamamız gerekenler şöyle özetlenebilir: N büyüdükçe kitle birimlerine ulaşmak maddi olarak ve de zaman kısıdı açısından tercih edilmez, örnekleme yapılarak parametre tahmini yapılır. Elbette mümkün tüm örneklemelerin çekilmesi söz konusu değildir, zaten böyle bir imkân olsa tüm kitleye ulaşılmış demektir ve tamsayım söz konusudur. Gerçekte, kitleye dair bir çıkarsama yapılacağı zaman bir tane örneklem çekilir. Ancak bu örneklemin çekilme yöntemleri ve hacminin ne kadar büyüklükte olması gerektiği hususları önemlidir. Bu konulara ileride ayrıntısıyla değineceğimizi belirterek, şimdilik basitçe şunları söylemekte yarar vardır.

Örneklemin kitleyi en iyi şekilde temsil etmesi gerekir, yani hanelerin aylık kira ödemeleriyle ilgili bir çalışma yapılacaksa, örneğin İstanbul için düşünersek, sadece Beşiktaş, Florya, Şişli'yi örneklem olarak seçersek İstanbul'u temsil etmiş olmayız. Ya da okulların başarı oranlarının hesaplanmak istendiği bir araştırmada sadece gelişmiş şehir merkezlerindeki okulları örnekleme alıp kırsal kesimi yok sayarsak bu durumda da gerçekten oldukça uzak, taraflı sonuçlar elde etmiş oluruz. Örneklemin temsili olmasının yanında, büyüklüğü de önemlidir. Örneğin 10 bin nüfuslu bir semtte araştırma yaparken 100 kişilik bir örneklem kullanmamalıyız. Çünkü bu rakam kitleyi temsil etmekten uzaktır.

Örneklem büyüklüğünün hesaplanması için çalışmanın başında belirlenen, araştırmacı tarafından kararlaştırılan belirli bir hata düzeyine göre örneklem hacmi hesaplanmaktadır. Basitçe ifade etmek gerekirse, büyük bir kitlenin görece büyük bir örneklemle temsil edilmesi gerekir. Örneğin 500 büyüklüğündeki bir kitle için 30 hacimli bir örneklem yeterli sayılabilirken, 10 bin hacimli bir kitle için örneğin 500 birimlik örneklem gerekebilir. Bu bağlamda, ülkemizde yapılan eğilim anketlerine dair açıklamaları değerlendirirken, ankete kaç kişinin katıldığını yani n örneklem hacmini göz önüne almak ve sonuçlara ancak yeterli örneklem üzerinden ulaşıldı ise güven duymak gerekmektedir. Türkiye'de çeşitli konulara ilişkin yapılan araştırmaların sonuçları medyada sık sık yer almaktadır. Fakat genellikle sadece bulgulara yönelik açıklama yapılırken, örneklemin kaç kişiden oluştuğuna değinilmemektedir.

Örneğin toplumun yarısının bayanlardan oluştuğunu düşünürsek bu rakam kabaca 40 milyondur. 20 milyonu çocuklardan oluşsa geriye 20 milyon yetişkin bayan kalır. Bu durumda Türkiye’deki kadınlara ilişkin bir araştırma yapılacağını düşünürsek, söz gelimi 500 kişi 1000 kişi gibi örneklemeler ile yapılan araştırmalara pek fazla itibar etmemek gerekmektedir. İstatistiksel araştırmanın temeli örneklemedir. Doğru, güvenilir ve yeterli veri yoksa en gelişmiş teknikler, modellemeler uygulansa dahi, kitleye dair doğru ve geçerli çıkarsama yapılmış olmaz. Gerçeğe yakın ve tutarlı sonuçlar elde etmek, doğrudan doğruya verinin kalitesine bağlıdır.

Birim:

Ölçülmeye ve (ya) sayılmaya elverişli olan canlı veya cansız tüm varlıklar istatistik birim olarak kabul edilebilir. Kişiler, hayvanlar, bitkiler, evler, arabalar, olaylar (evlenme, boşanma, ölüm, depresyon, trafik kazaları, seller, çığ düşmeleri) v.b. hakkında ölçüm veya sayım yapılabileceği için istatistiksel birimdirler. Ancak, korku, hayal, rüya, v.b. ölçülemeyen, sayılamayan soyut varlıklar istatistiksel birim değildirler ve istatistiksel çalışmalara konu olamazlar.

Birimler 3 grupta incelenebilirler:¹⁸

Sürekli olanlar <i>İnsan, bina, şirket, v.b.</i>	Doğal olanlar <i>İnsan, kitap, v.b.</i>	Gerçek olanlar <i>Ev, arsa, bisiklet, v.b.</i>
Ani olanlar <i>doğum, evlenme, boşanma, v.b.</i>	Doğal olmayanlar <i>arsa, pasta, v.b.</i>	Varsayıma dayananlar <i>örneklem</i>

Değişken:

Birimlerin araştırmaya konu olan herhangi bir özelliğine değişken denir. Birimlerimiz her bir öğrenci olduğunda bunların boy ölçüleri, ağırlık ölçüleri birer değişkendir, öğrenciden öğrenciye değişir. Ailelerin gelir düzeyleri, sahip oldukları çocuk sayısı bir değişkendir. Değişkenlerin alabilecekleri her bir değere ise “şık” denir. Örneğin medeni hal bir değişken olarak ele alındığında evli, bekâr, dul, vb. değerlerinin her biri birer şıktır.

Değişkenler;

a. Sayısal (nicel, kantitatif) değişkenler

Boy, ağırlık, alan, hacim, vb.

b. Sayısal olmayan (nitel, kalitatif) değişkenler

¹⁸ (Murat Açıköğretim Yayınları, ss: 1-2)

Cinsiyet, meslek, göz rengi, memleketi, vb. olarak ikiye ayrılır.

Sayısal değişkenler de;

a.Sürekli

b.Kesikli

olmak üzere ikiye ayrılır. Sürekli değişkenler ölçmeye tabidirler. Kesikli değişkenlerde ise sayma söz konusudur. Ailedeki çocuk sayısı, bir caddedeki evlerin sayısı, bir öğrencinin sahip olduğu kitapların sayısı, vb. bunlar hep sayılabilen yani kesikli değişkenlerdir. Değişkendir çünkü kişiden kişiye, aileden aileye farklı değerler almaktadır. Sürekli değişkenlere örnek verecek olursak, bir kutunun ağırlığı, bir insanın boy uzunluğu, bir arazinin yüz ölçümü, vb. sürekli değişkenlerdir. **Özetle, eğer sayısal bir değişken sayılabiliyorsa kesiklidir, ölçülebiliyorsa sürekli.**

Nitel veriler kendi aralarında ikiye ayrılırlar; nominal veri ve ordinal veri.

Nominal veri; kategori ifade eder, Kategoriler arasında bir sıralama veya üstünlük yoktur. İstatistiksel araştırmalara sayısal kodlama yapılarak sayısallaştırılmak suretiyle dâhil edilirler. Örneğin;

Medeni hal: evli (1), bekâr (2), ayrılmış (3)

Göz rengi: siyah (1), yeşil (2), kahverengi (3),vb.

Cinsiyet: Kadın (1), erkek (2)

Meslek: öğretmen (1), bankacı (2), avukat (3), sanatçı (4), vb.

Ordinal veri; verinin sıralanması şeklindedir. Kategoriler arasında bir üstünlük sıralaması söz konusudur.

Örneğin;

Zayıf (1), orta (2), iyi (3), çok iyi (4)

Eğitim durumu: ilkököl (1), ortaokul (2), lise (3), üniversite (4), yüksek lisans (5), doktora (6)

Örnek:

a) Ekonometri bölümünde okuyan öğrencilerin evlerinden fakültelerine olan uzaklıklar

Bir ölçüm söz konusu olduğu için sürekli nicel veridir.

b) Bir okuldaki öğretmenlerin cinsiyetleri

Kategori belirten, nominal veridir.

- c) 2014 yılında İstanbul'a düşen kar yağış miktarı
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
- d) Son 5 yılda Tunceli-Erzincan karayoluna düşen çığ sayısı
Sayma işlemine dayandığı için kesikli nicel veridir.
- e) Göz rengi
Kategori belirten, nominal veridir.
- f) Bir evin tavan yüksekliği
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
- g) Bir ilacın kana karışma süresi
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
- h) Mezun olunan okul, askeri rütbelere, akademik ünvanlar
Sıralama söz konusu, ordinal veridir.

2.3. İstatistikte Hata Kavramı

Genel olarak iki tür istatistik hatadan bahsedilebilir;

1. Rastsal (tesadüfi) hata
2. Sistematiik hata

Rastsal hata adı üzerinde, bilinçsizce yani herhangi bir kasıt olmadan yapılan hatalardır. Bu tür hatalar örneklem büyüdükçe artı ve eksi yönde birbirini götürerek etkisiz hale gelecektir. Bu nedenle rastsal hatalar genellikle göz ardı edilirler. Örneğin bir anket uygulamasında kişinin cinsiyeti “kadın” olduğu halde “erkek” şeklinde giriş yapılmış olabilir. Aynı anketör bir başka anket formunda tam tersi bir işaretleme yapmış olabilir veya bir başka anketör onun yaptığının tersi bir veri girişi yapmış olabilir. Yani bu ve benzeri bazen bir yönde bazen de tersi yönde yapılan hatalar **gözlem sayısı arttıkça denkleşerek toplamda göz ardı edilebilmektedirler**, böyle hatalara rastsal (tesadüfi) hata denir ve sakıncası yok denecek kadar azdır. 16. yy’da Bernoulli ve diğer matematikçiler tarafından yapılan “büyük sayılar kanunu”, gözlem sayısı arttıkça olayların rastsal nedenlerin etkisinden kurtulduğunu ortaya koymaktadırlar.

Sistematiik hata ise bazen kasıtlı olarak yapılan bazen ise kullanılan teçhizatın bozukluğundan kaynaklanabilen, hep aynı yönde seyreden ve örneklem büyüse de tolere edilemeyen hatta bu şekilde daha da artan yapıdaki hatalardır. Sonuçların objektifliğini etkileyeceği için **kesinlikle düzeltilmeleri gerekir**. Örneğin hükümet karşıtı bir anket uygularken soruları hep muhaliflere yöneltmek sonuçların yanlı ve tek yönlü çıkmasına neden olacaktır. Sistematiik hatalar bazen de kullanılan teçhizatın bozukluğundan kaynaklanabilir. Örneğin bozuk bir klavyenin “h” harfini devamlı olarak “g” olarak basması, bir tartının ayarının bozuk olması nedeniyle hep 15 gr eksik tartması, vb. Verilen örneklerden de anlaşılacağı üzere, sistematiik hatalar yapılan istatistiksel çalışmanın en başından kalitesiz olmasına neden olur. İstatistiksel çalışmaların en temel ihtiyacı öncelikle doğru ve sağlam verilerdir.