

Bölüm I

İstatistik

Gün gelecek; istatistik, etkin vatandaşlık için okuma yazma kadar gerekli olacak. (Herbert G. Wells)

Üç türlü yalan vardır: Yalan, kuyruklu yalan ve İstatistik... (Benjamin Disraeli)

İstatistik; güvenilmez gerçeklere güvenilir rakamlar üretme bilimidir. (Evan Esar)

İstatistik dersi grip açısından daha sevimli değildir ama en az o kadar yararlıdır. (Paul Neubold)

İstatistik laneti insanlığı sarmadan önce mutlu, masum, neşe dolu bir hayat sürer, hayli iyi hükümlere varırdık. (Hilaire Belloc)

Neyi söylemediğini dikkatle değerlendirene dek, istatistiklerin söylediği şeye güvenme. (William W. Watt)

İstatistiklerle yalan söylemek kolaydır. Ama onlarsız doğruyu söylemek çok daha zordur. (Charles Frederick Mosteller)

İstatistikler herhangi bir şeyi – hatta gerçeği – kanıtlamak için yapılmış olabilirler. (Noel Moynihan)

İyi ambalajlanmış bir istatistik, Hitler'in büyük yalanından daha etkilidir. (Darell Huff)

Bütün bilgiler son tahlilde (analizde) tarihtir. Bütün bilimler teorik olarak matematiktir. Bütün hükümler mantıksal açıdan (mantıksal olarak) istatistiktir. (C.R. Rao)

Önceki zamanlarda istatistik yoktu, bu yüzden hep yalanlarla boşuşmak zorunda kalmışlardı. (Stephen Leacock)

İstatistik, sayısal yöntemlerin zaferidir; sayısal yöntemler ise verimsizliğin ve ölümün zaferidir. (Hilarie Belloc)

İstatistik ezikler içindir. (Scotty Bowman)

İstatistikçi, matematikçi gibidir. Ne hakkında konuştuğunu bilir ve söylediği her şey matematiksel olarak doğru olabilir. Fakat doğru bir karar veremez. (W.P. Coole)

İstatistikçi, rakamları toplayan ve onları yoldan çıkaran kişidir. (Evan Esar)

Son yüzyılın gidişatı, istatistikçilerin lehinedir. (H.G. Wells)

1900’lü yılların başında, ünlü İngiliz yazar Herbert George Wells “Gün gelecek, istatistik, etkin vatandaşlık için okuma yazma kadar gerekli olacak” sözü ile sanki günümüzü tarif etmiştir. Değişen Dünya koşulları ve ihtiyaçlar, ilerleyen teknoloji, kalabalık insan nüfusu hemen her konuda daha iyiyi aramamıza etken olmaktadır. Günümüzde her alanda rekabet artmış durumdadır. İster bireysel ister kurumsal olsun, 21. yüzyılda rekabet etmek istiyorsanız, yeniliklere açık olmalısınız. Bu yenilikler teknolojik, dilbilimsel veya sayısal olarak karşımıza çıkabilir. Önemli olan geçmiş, günümüz ve gelecek arasındaki köprüyü başarılı bir şekilde kurmaktır. İstatistik, bizlere bu köprüyü kurmamızdaki en önemli katkıyı sağlayan bilim dallarından bir tanesidir.

Temel amaçlarının arasında durum tespiti yapmak, karşılaştırmak, tahmin etmek ve karar vermek olan istatistik, hemen her alanda kullanılmaktadır. İstatistik için birçok tanım yapılmıştır. Her bir tanım, istatistiği farklı açılardan ele alıp tanımlamaya yardımcı olmuştur. Tek bir tanım yerine birçok tanımın anlatmak istedikleri ile istatistik dünyasına giriş yapalım:

İstatistik, bir sonuç çıkarmak için verileri yöntemli bir biçimde toplayıp sayı olarak belirtme işidir veya sayımlama yapmaktır.

İstatistik, basit anlatımla veri toplama ve bunları yorumlama bilimidir.

İstatistik, geçmiş ve şimdiki durum hakkında toplanmış sayısal verileri, geliştirilmiş olan bazı matematiksel tekniklerle analiz ederek gelecek hakkında karar vermemizi kolaylaştıran bir bilim dalıdır.

İstatistik, bir popülasyonu tanımlamak için ilgili popülasyondan belirli yöntemlerle elde edilen örnek verilerin analizine dayanarak popülasyon ile ilgili tahminler yapan ve sonuç olarak yorumlarda bulunan bir bilimdir.

İstatistik, herhangi bir gerçekliğin veya olgunun sayısal doğasını anlamamıza ve başkalarına anlatmaya yarayan bir bilim dalıdır.

İstatistik, çok sayıda dış etkene bağlı nesne, varlık ya da olayların sayısal dökümü yapılabilen özelliklerini, incelemeye yarayan bir teknik ya da yöntem kümesidir.

İstatistik, sayısal bilgiler arasında karşılaştırmalar yaparak incelenen özelliklerin birbirleri ile ilişkilerini belirleyerek geleceğe yönelik tahminlerde bulunmak üzere, sayısal verilerin analiz edilerek yorumlanmasına yardımcı bir bilim dalıdır.

İstatistik, verilerin toplanıp organize edildiği ve sonuçların yorumlandığı bilim dalıdır.

İstatistik, ilgilendiğimiz olayların gözlenerek ilgili verilerin derlenmesi, işlenmesi, analizi ve yorumlanmasında kullanılan tekniklerin tümünü ifade eder.

İstatistik, belirli amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan derlenen sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimidir.

İstatistik, bir yandan resmin bütününe görmeye, göstermeye uğraşır, diğer yandan tekil gözlemleri de göz ardı etmez.

İstatistik biliminin hemen her alanda kullanıldığından bahsetmiştik. İstatistiğin kullanımının sağladığı yararları kısaca şu şekilde özetleyebiliriz:

1. Araştırmacılara, elde ettikleri bulguları açıklarken yardımcı olur.
2. Araştırmacıların, verilerin yardımıyla genel sonuçlara ulaşmasına yardımcı olur.
3. Anakitleye ulaşamadığı durumlarda zaman ve maliyet tasarrufu sağlar.
4. Çevremizi daha iyi tanımamıza (sadece görsel olarak değil sayısal olarak da) yardımcı olur.

İstatistiğin saydığımız yararları dışında en önemli yararı, disiplinler arası bir bilim dalı olması nedeniyle, birçok bilim dalı ile çalışabilmesidir. İktisadi olayların gelişimini, nedenlerini incelemek için iktisat bilim dalıyla istatistik bilim dallarının birleşimi olan ekonometri bilim dalı buna en iyi örnektir. Benzer şekilde kimya bilim dalı ile kemometri, psikoloji bilim dalı ile psikometri, sosyoloji bilim dalı ile sosyometri gibi bilim dalları ile yeni bilim dalları geliştirilmiştir.

İstatistik tanımlamasını yaparken, çok sık kullanılan bir deyiş karşımıza çıkar: *Rakamlar yalan söylemez, ama yalancılar rakam söyler*. İstatistiğin kullanımını ve gücü hakkında basit bir latifedir aslında bu söz. Yine de gerçeklik payı taşıması bile, bizlere istatistik ve rakamlar hakkında bir soru işareti bırakmaktadır. Her araç iyiye de kullanılır, kötüye de. Elektrikle ister aydınlanma yaparsınız, ister işkence. Benzer şekilde istatistik ile doğruları gösterebileceğimiz gibi insanları aldatabiliriz de.

Olayları sayısal değerlere dayandırarak, işimize geldiği taraftan bakma alışkanlığı ile gerçekleri saptırma alışkanlıklarımız vardır. Böylece istatistik (dolayısıyla bilim), yalan söyleme aracı olarak kullanılabilir. Bu yanılgıdan kaçınmak için istatistiği iyi öğrenmeli, rakamların ve istatistiksel bilgilerin ardındaki gerçekleri çok iyi kavramalıyız. Genel olarak istatistiği kötüye kullanma biçimlerini şu şekilde sıralayabiliriz:

- i. Konuyu bilmemek

- ii. Hatalı verilerle çalışma yapmak
- iii. Araştırmayı yanlış planlamak
- iv. Sonuçları yanlış ya da yanıltıcı yorumlamak
- v. Yanlış ya da yetersiz çözüm yöntemleri kullanmak
- vi. Sonuçları çarpıtmak

Yukarıda sayılan maddeler dışında da birçok kötüye kullanım biçimi sayabiliriz. Önemli olan, bu hatalardan mümkün olduğunca kaçınmak olmalıdır. Bunun için de uzman tavsiyesi almak, kaynak kitaplardan yararlanmak veya iyi araştırma yapmak gerekmektedir.

İyi bir istatistik bilgisi, araştırma için yeterli olmayabilir. Bunun için bazı aşamaların uygulanması, çalışmanın başarısı açısından önem göstermektedir. İstatistiksel bir araştırmada şu aşamalardan söz edebiliriz.

1. *Amacın belirlenmesi:* Araştırmanın amacı; var olan bir teorinin, bir gerçekliğin eleştirilmesi veya bütünüyle bir teorinin reddedilmesi ve yeni bir teorinin öne sürülmesi olabilir.
2. *Verinin toplanması:* İlgili veri toplama araçları yardımıyla veriler toplanarak istatistiksel araştırma için uygun hale getirilmelidir. Veriler; resmi kaynaklardan elde edilebileceği gibi (TÜİK, Merkez Bankası vs.) anket yoluyla, gözlem yoluyla veya deney sonucunda da elde edilebilir.
3. *Toplanan verilerin işlenmesi ve düzenlenmesi:* Düzenlenen veriler tablo veya grafikler ile gösterilerek durum değerlendirmesi yapılabilir.
4. *İstatistiksel analiz ve tahmin:* Analiz ve tahmin aşamasında, önceden belirlenen istatistiksel yöntemler kullanılarak, ilgili veriye analizler yapılır.
5. *Karar ve yorum:* Analiz yapıldıktan sonra ortaya çıkacak sonuçlar değerlendirilerek yorumlanır.

Her gün aynı sayıda satış yapan bir satıcı veya aynı sayıda matematik sorusu çözen bir öğrenci için, istatistik önemli olmayacaktır. Çünkü bir sonraki gün veya adımda ne olacağını tahmin etmek zor değildir. Bir önceki günün aynısı olacaktır. Fakat her gün değişen sayıda satış yapan bir satıcı için her yeni gün bir soru işaretidir. Bir önceki günün veya haftanın satışlarının önemi bile değişecektir. İşte bu tip birbirinin tam benzeri olan olaylara *tipik olaylar* denir. Bu durumda çok sayıda olayı incelemeye gerek kalmayacaktır. Birbirlerinin aynısı olduğundan tek bir olayı inceleyerek kolayca yorum yapabiliriz.

Benzer şekilde, birbirine benzemeyen fakat bazı ortak noktaları olan olaylara *kollektif olaylar* denir. Olaylar birbirlerine tam olarak benzemediği için bu olaylardan herhangi biri, oluşturdukları topluluğu temsil edemeyecektir. Bu konu için en iyi örnek insanlardır. Birbirlerine tam olarak benzemeseler de (ikiz bile olsalar, farklı huya veya alışkanlıklara sahip olabilirler) insan olmanın verdiği birçok ortak özelliği de barındırırlar. Kollektif olayları oluşturan etkenlerin araştırılması, istatistiğin birinci aşamasıdır aslında. Bu etkenlerin çoğunun rastlantısal olması, genelleme yapmayı da zorlaştırmaktadır. Genel nedenlerden çok rastlantısal etkenlerden etkilenen olaylar da birbirlerinden farklı olurlar. Böylece genel düzenin de dışına çıkmış olacaktır. İşte bu etkenlerin, ilgilenilen olaylara etkisi araştırılmak istendiğinde istatistiğe başvurulmalıdır. Sıradanlığın dışında gerçekleşen her türlü davranış, ölçme veya sayma işlemini araştırmak, incelemek ve yorumlamak, istatistiğin konusunu oluşturmaktadır.

Bölüm 2

Temel Kavramlar

Birçok insan size yanlış bilgi verir. (Edmund C. Berkeley)

Bir kişi ne olması gerektiğini düşünmelidir, düşündüğü şeyin ne olmasını değil. (Albert Einstein)

Teori, gözlüklerin buğulanması gibidir. Belirsiz gerçeklerdir. (Charlie Chan)

Birinci sınıf teori işleri tahmin eder, ikinci sınıf teori işleri yasaklar, üçüncü sınıf teori ise işleri açıklar. (Aleksander I. Kitaigorodski)

Keşke yaptığı açıklamayı açıklayabilseydi. (Lord Byron)

Gerçekler, istatistiklerden daha yüksek sesle konuşurlar. (Justice Streetfield)

İstatistiksel bir ilişki ne denli güçlü ve ne denli anlamlı olursa olsun, asla nedensel bir ilişki kuramaz. Bizim nedensellik düşüncelerimiz istatistiğin dışından, eninde sonunda şu ya da bu kuramdan gelmelidir. (M.G.Kendal - A.Stuart)

Bu bölümde, istatistik konularında yer alan temel kavramlara kısaca değinilecektir. Kavramların iyi tanımlanması ve anlaşılması, ilerideki konuların da iyi anlaşılması açısından önemlidir.

Bilgi: Bilgi, bir deneyim veya eğitim sırasında bir kişinin ihtiyaç duyduğu uzmanlık ve yetenekler olarak özetlenebilir.

Olay: Bir deneyin sonucu olay olarak tanımlanır. Örneklem uzayındaki herhangi bir alt küme bir olaydır. Bir zar atma deneyinde 3 gelmesi, bir futbol takımının bir turnuvada 4 galibiyet ve 2 beraberlik alması, istatistik vize sınavından 75 alınması birer olaydır.

Tek bir sonucu olan olaya basit olay, birden fazla sonucu içeren olaylara da bileşik olay adı verilir. Bir tavla zarının atılması ile 5 sayısının ortaya çıkması basit olaydır. İki zarın birlikte atılmasında toplamın 3 olması için çıktılar (1,2) ve (2,1) durumları olup, bu bir bileşik olaydır.

Birim: Bütünü oluşturan her türlü olaya birim adı verilir. Diğer bir ifade ile hakkında ayrı ayrı bilgi toplanan bir bütünün (anakitlenin) en küçük parçasıdır. İnsan, bilgisayar, banka şubesi, ders kitabı, müzik albümü, ölüm kavramları birer birim olarak tanımlanabilir. Birimler, *sayılabilir* veya *ölçülebilir* olmalıdır. Zevkler, tercihler veya rüya gibi kavramlar birim tanımına girmezler.

Belirlenen amaç ya da amaçlar doğrultusunda, ilgilenilen yığın olayın tanımlanması ile kimlerin ya da nelerin gözleneceği belirlenmelidir. Ancak birim seçilirken, amaca uygunluk ve uygulanabilirlik özelliklerinin öncelikle göz önünde bulundurulması gerekir. Bunun için de birim belirlenirken, birim tanımının kesin, amaca uygun ve uygulamaya elverişli olması gerekmektedir.

Gözlem: Birimlerde incelenen özelliklerin, gözlenerek veya ölçülerek elde edilen değerleridir. Gözlem; birimlerin, her türlü özelliklerinin belirtilme işlemidir.

Veri: Gözlem değerlerinin sayımlarından ya da ölçümlerinden oluşan *istatistiksel veri* ya da kısaca *veri*; araştırmaların ham maddesini oluşturur. Toplanan verinin henüz düzenlenmemiş haline *ham veri* denir. Veri; genel nüfus sayımında olduğu gibi sadece betimsel amaçlarla toplanabilir veya genel seçim öncesi bir bölgenin oy oranını tahmin etmek için anket sonuçları da olabilir. Amaç ne olursa olsun verinin yapısının ayrıntılarıyla ortaya çıkartılması, özlü ve yanıltıcı bir amaç güdülmeden sunulması çok önemlidir.

Anakitle (Yığın, Anakütle, Popülasyon, Evren): Üzerinde araştırma yapılan herhangi bir canlılar veya cansızlar topluluğuna anakitle denir. Cinsiyet, bir üniversitede okuyan öğrenciler, futbolda Türkiye 1. ligi takımları, dünya ülkeleri birer anakitle olabilirler. Türkiye’de üniversitelerde okuyan yaklaşık 6

milyon öğrenci vardır. Eğer tüm üniversite öğrencileri ile ilgili bir çalışma yapıyorsak, anakitlemiz 6 milyon öğrencinin hepsi olacaktır. Fakat araştırma konumuz sadece Cumhuriyet Üniversitesi Ekonometri bölümü öğrencileri ile ilgiliyse, anakitlemiz bu bölümde okuyan 500 öğrenci ile sınırlı olacaktır.

Tamsayım: Anakitlede bulunan tüm birimlerden veri toplanırsa, tamsayım olarak tanımlanır. Tamsayım sonuçları, kitle hakkında en doğru ve kesin bilgiyi verir. Nüfus sayımı, tamsayım için ideal bir örnektir.

Örnek (Örneklem): Anakitlenin içinden çekilen birimlerden oluşan alt topluluğa *örnek (örneklem)* denir. Tamsayım yapmak çoğu zaman zordur. Maliyet, zaman, çalışacak eleman bulma gibi zorluklar tamsayım yapılmasını zorlaştırır. Bunun yerine kitleden belirli bir sayıda birim üzerinde araştırma yapmak ve istatistiksel teknikler ile kitle hakkında tahminlerde bulunmak, zaman ve maliyet gibi konularda tasarruf sağlamamızı sağlar.

Örnekleme: Anakitleden örneklem seçme işlemine örnekleme denir. Anakitle çok büyük olduğu zamanlarda veya tüm anakitle elemanları hakkında bilgiler elde etmek imkânsız olduğu durumlarda, anakitleyi değiştirmek veya elemanlarına ulaşmak çok büyük masraf ve emek gerektirecektir. Bu nedenle tamsayım ile tüm sayısal değerlerin öğrenilmesi istenmez. Örnekleme ile kullanılabilir bir büyüklükte altküme elde edilir. Örnek verileri toplanır; örnek verileri hakkında istatistikler hesaplanır ve örnek istatistikleri üzerinde çıkarımsal istatistikler yapılabilir. Milletvekili seçimleri öncesi piyasa araştırmaları, örnekleme konusundan yararlanırlar.

Örneklem Büyüklüğü: Örneklem büyüklüğü, çalışılan örneklemdeki toplam gözlem, birey ya da ölçüm sayısıdır. Örneklem büyüklüğü genellikle N ile gösterilir. Eğer örnekleme 45 kişi varsa, örneklem büyüklüğü $N=45$ 'tir.

Örneklem Uzayı: Herhangi bir deneyin mümkün olan tüm sonuçlarının oluşturduğu kümedir. Örneklem uzayı S ile gösterilir. Bir zar atma deneyinde, zarın üst yüzeyine gelen sayının tüm mümkün sonuçları örnek uzayını oluşturacaktır ve $S = \{1, 2, 3, 4, 5, 6\}$ ile gösterilir.

Parametre: Anakitleyi temsil yeteneğine sahip karakteristik değerlere parametre denir. Genelde bilinmeyen ve tahmin edilmeye çalışılan bir değerdir.

Değişken: Bir şahsın, grubun veya çevrenin bir durumdan diğerine değişebilen veya bir farklılık gösterebilen herhangi bir özelliğidir. Başka bir tanımla, farklı birimler için farklı değerler alabilen ve incelenen tüm özelliklere değişken adı verilir. Cinsiyet, ağırlık, mesleki saygınlık, politik ideoloji, ırk, maden çeşidi gibi örnekler değişken olarak adlandırılabilir.

Değişkenler farklı kullanım alanlarına göre farklı isimlendirilirler.

i. Nitel - Nicel Değişken

Değişkenler sözcüklerle ifade edilirse *nitel değişken*, sayılarla ifade edilirse *nicel değişken* adını alırlar.

ii. Sürekli –Kesikli Değişken

Sayma veya ölçme suretiyle ifade edilen değişkenler *sürekli değişken*, sayma veya ölçme suretiyle ifade edilemeyen değişkenler ise *kesikli değişken* olarak adlandırılır.

İşlem: İşlem, araştırmacının deneysel bir araştırmada deneysel birimlere uyguladığı her bir faaliyettir. Örneğin, bir öğretmen farklı gruplara farklı öğretim metotlarını uygulayarak daha iyi sonuç almaya çalışabilir veya bir doktor farklı ilaçların baş ağrısı üzerindeki etkilerini incelemek üzere bir çalışma gerçekleştirebilir. İşlemler deneysel birimlere miktar veya büyüklük ile açıklanabilen düzeyler ile uygulanır. Bir kimyasalın A, B ve C alt türlerinin etkisi inceleniyorsa, işlem üç düzeylidir. Bir öğretim görevlisi, gece ve gündüz öğretim gören öğrencilerin, İngilizce ve Türkçe istatistik derslerini karşılaştırmak istiyorsa, iki grup öğrenci için iki düzeyli bir işlem gerçekleştirmelidir.

Oran: Belirli bir özelliğe sahip birim sayısının, toplam birim sayısına bölünmesi ile bulunur. Örneğin, 10 tane topun 6'sı kırmızı renkte ise, kırmızı renk topların oranı $6/10 = 0.6$ 'dır. Araştırma çalışmalarında oran, belli bir özelliğe sahip olan deneklerin frekansının örneklem büyüklüğüne bölünmesi ile bulunur.

İstatistik kullanım amacına göre ikiye ayrılır:

i. Betimsel (Açıklayıcı) İstatistik

İlgilenilen anakitledeki tüm birimlerden ilgili değişken veya değişkenler için veri toplandığında, bunları kullanarak anakitle hakkında kullanışlı özet bilgiler elde etmek amacıyla toplanan istatistiksel bilgilerdir.

Betimsel istatistik ile verileri düzenleyebilir, tablolar ve grafikler ile özetleyebiliriz. Verilerin ortalama ve değişkenlik ölçülerini bularak, verinin bütününe özetleyen istatistiksel kavramları hesaplayabiliriz.

ii. Çıkarımsal İstatistik

Anakitleden rastgele çekilen örneklemde toplanan verileri kullanarak, anakitle hakkında tahminler yapmayı veya parametreler hakkında hipotezlerin doğruluğunu amaçlayan istatistiksel testler veya çalışmalardır.

Çıkarımsal istatistik ile bütününe ulaşamayan anakitle hakkında örneklem seçebilir, bu örneklem bilgilerine dayanarak anakitle hakkında çıkarımlar ve tahminler yapabiliriz.

Bölüm 3

Verilerin Düzenlenmesi

Hiçbir zaman bir şeyler anlatmaya çalışmayın. Onu şekillerle ifade etmeye çalışın. Kafamızla değil, gözlerimizle öğreniriz. (Will Rogers)

Bir kroki, bir düzine sayfadan daha çok şey anlatır bana. (Ivan Turgenev)

Herkes, iyi bir veriyi kolayca kötü kullanabilir. (William E. Demings)

Veri olmadan teori üretmek büyük bir hatadır. (Sherlock Holmes)

Hiçbir insan zihni, sayısal verilerin önemli miktarlardaki anlamını kavramaya yetmez. (Ronald A. Fischer)

Verileri dışlamanın mantığını anlamıyorum, onlar zaten inanılmaz görünüyorlar. (Fred Hoyle)

Model kurmanın amacı, verilere uygun olsun diye değil, problemi kuvvetlendirmek içindir. (Samuel Karlin)

Veri toplandıktan sonra istatistikçi çağırarak, cenazeyi morga gittikten sonra doktoru çağırarak gibidir. O size ancak neyin yanlış gitmiş olduğunu söyleyebilir, ama yanlış düzeltilmesi artık pek mümkün değildir. (Ronald Fisher Box)

Veri, veri, veri... Çamur olmadan tuğla yapamam ki... (Sherlock Holmes)

Ölçün, ölçemiyorsanız ölçmeye çalışın. (Galilei Galileo)

İyi bir grafik, bizi asla görmeyi beklemediğimizi görmeye zorlar. (John Tukey)

3.1. Veri

Gözlem değerlerinin sayımlarından ya da ölçümlerinden oluşan *veri*; araştırmaların ham maddesini oluşturduğundan bahsetmiştik. İstatistiksel analizler yapabilmek için verilere ihtiyaç duyarız. Her çeşit veri ile istatistiksel çalışmalar yapılabilir. Ancak çalışmaların güvenilir olabilmesi için verilerin de güvenilir olması gerekmektedir. Veriler çeşitli yollarla elde edilebilir. Resmi kurumların yayınladıkları veriler, geniş kitlelerin sonuçlarını da içerdiklerinden en kapsamlı olanlarıdır. Bunun yanı sıra gözlem, ölçüm veya deney yoluyla kendi verilerimizi de toplayabiliriz. Toplanan verilerin büyüklüğü, verilerin derlenmesini ve düzenlenmesini de gerektirir.

Veri derlenirken sadece belirlenen amaçlar doğrultusundaki değişkenler göz önünde tutulmalıdır. Uygulamalarda fazla ayrıntı, sorunlara neden olabilir. Öte yandan uygulamalarda gözlem sayısı kesinlikle sonlu bir sayı olacaktır. Ayrıca, gözlemlere bağlı olarak ilgili değişkenlerin alacakları değerler de (şıklar da) ilgili değişken sürekli ya da süreksiz olsun, sonlu olacaktır.

Şıklar belirlenirken, dikkat edilmesi gereken önemli bir nokta da gözlemlerde kullanılan ölçü biriminin, araştırmanın doğasına uygun olması gereğidir. Örneğin; ayçiçeği yağı üreten bir firmanın, aylık üretimi için kilo, şişe, teneke ya da ton makul ölçüler iken, bir kuyumcunun bir günde sattığı bilezikler için gram uygun bir ölçü olmaktadır.

3.1.2. Veri Derleme Çeşitleri

Veri derleme, kullanım amaçlarına ve derleme sürelerine göre çeşitli durumlara ayrılmaktadır.

Ani - Sürekli Derleme

Eğer gözlemlenecek kitledeki birimler sürekli karakterdeyse, istenilen bir anda gözlenmeye hazır olan bu tür birimlerin gözlenmesi ya da kaydedilmesi işlemlerine *ani veri derleme* denir. Nüfus sayımları, seçim anketleri ve iş yeri sayımları bu tür veri derlemeye örnek oluşturur.

Eğer ilgilenilen kitle ani birimlerden oluşmuşsa (bu tür birimler zamana yayıldığından), belli bir zaman aralığında gözlenmeleri ve kaydedilmeleri gerekir. Bu tür işlemlere *sürekli veri derleme* denir. Belirli bir bölgede ve zaman aralığında evlenmeler, boşanmalar, trafik kazaları, doğumlar ve ölümlere ilişkin derlenen veriler, bu tür veri derlemeye örnek oluşturur.

Genel - Kısmi Derleme

Hakkında bilgi edinilmek istenen kitlenin tamamının gözlenmesine *genel veri derleme* adı verilir. Genel nüfus sayımı ve tarım rekolteleri birer genel veri derlemedir.

Hakkında bilgi edinilmek istenen kitleyi oluşturan birimler arasından, belirlenen amaçlar doğrultusunda yalnızca bir kısmının seçilip gözlenmesine, *kısmi veri derleme* adı verilir. Kısmi veri derleme, genel veri derlemenin pahalı oluşu, zaman alışı, gözlem birimlerinin fiziksel zarara uğraması gibi nedenlerle yapılmak istenmediği zaman uygulanır. Seçim anketleri genellikle bu yöntem ile yapılır. İstanbul'da yaşayan 14 milyon kişi ile anket yapmak yerine çok daha az kişi ile anket yaparak (5000 kişi ile), elde edilen sonuçlar genele göre yorumlanır.

Dolaylı – Dolaysız Derleme

Eğer bir anakitle hakkında bilgi edinmek için sadece o anakitle gözlemleniyor ise *dolaysız derleme* söz konusudur. İlgilendiğimiz anakitle hedefimizdir ve başka bir şey ile ilgilenmeyiz. Fakat bilgi edinmek istediğimiz anakitle hakkında başka anakitle birimlerinin gözlenmesi durumunda *dolaylı derleme* adını alır.

Birincil – İkincil Derleme

İstatistik amaçlar ile yaptığımız derlemelere *birincil derlemeler*, diğer amaçlar (idari, mali, siyasi vs.) için yaptığımız derlemelere *ikincil derlemeler* adı verilir. Suç sayıları birincil derleme olurken, işlenen suçların türlerine göre ayrıştırılıp dava sürecinde kullanılması ikincil derlemedir. İkincil derlemeler çoğunlukla birimlerin kendi ihtiyaçlarından doğmaktadır.

Asıl – Yardımcı Derleme

Anakitle birimlerinin özelliklerinin belirlenmesi için yapılan derlemeye *asıl derleme* adı verilir. Asıl derlemeye yardımcı olması istenen derlemeler ise *yardımcı derlemeler* olarak adlandırılır. Yardımcı derlemelere daha çok kontrol amaçlı başvurulur.

3.1.3. Veri Derleme Araçları

Veri derleme işlemleri Türkiye İstatistik Kurumu, Merkez Bankası, DPT gibi kamu kuruluşları ile PİAR, GALLUP gibi kar amacı güden özel kuruluşlarca da yapılır. Bu kurumlar, veri toplamak için konu ile ilgili ölçüm veya deney cihazları kullanırlar. Bu araçların kullanılması istatistik alanının dışında kabul edildiğinden, burada istatistiksel veri toplamada kullanılan veri derleme araçlarına kısaca değinilecektir.

Soru Kâğıtları

Soruları içeren ve cevapların üzerine yazılacağı kâğıtlara genel bir ifade ile *soru kâğıdı (bülten)* adı verilir. Eğer soru kâğıtlarında; tek birime ilişkin sorular yer alıyorsa *bireysel bülten*, çeşitli birimlere ilişkin sorular yer alıyorsa *kollektif bülten* adını alırlar.

Gözlemciler

Veri derlemelerde, birimleri ve özelliklerini kaydeden kişilere genel anlamda *gözlemci* veya *görüşmeci* denir. Gözlemcilerin; verilen cevapları doğru ve eksiksiz bir şekilde kayıt etmeleri araştırmanın güvenilirliğini de doğrudan etkilemesinden dolayı, gözlemciler istatistiksel çalışmalarda önemli bir yere sahiptirler.

3.1.4. Veri Derlemenin Uygulanması

Veri derlemenin uygulanması aşamasında, incelenecek anakitle veya örneklemdeki birimler hakkında bilgi toplanmaktadır. Uygulama genelde üç şekilde yapılmaktadır: Gözlem, görüşme ve anket. *Gözlem*, olayları sistematik bir biçimde incelemektir. *Görüşme*, sözlü iletişim (karşılıklı konuşma) yoluyla veri toplama tekniğidir. *Anket* ise yazışma, yazılı iletişim (karşılıklı yazı yazma) yoluyla veri toplama tekniğidir. Bu üç teknik de uygulaması kolay tekniklerdir. Bu teknikler, dikkatli ve sistematik uygulandıklarında son derece başarılı sonuçlar elde edilir.

3.1.5. Veri Derleme Hataları

Veri toplama ve derleme çalışmalarında ne kadar dikkatli olsak da hatalar olması muhtemeldir. Bu hataların kaynaklarının bilinmesi veya fark edilmesi, olası yanlış sonuçların da önüne geçilmesini sağlayabilir. Genel olarak hataları şu şekilde sıralayabiliriz:

1. *Derlemeyi düzenleyenlerden kaynaklanan hatalar:* Derlemeyi düzenleyenlerin yanlış zamanda araştırma yapmaları, çevresel ve etnik faktörlerin dikkate alınmaması, yetersiz eleman çalıştırma veya yanlış yöntem kullanmaları bu tip hataların kaynakları arasında sayılabilir.
2. *Cevaplayıcılardan kaynaklanan hatalar:* Özellikle anket çalışmalarında, cevaplayıcı beğenmediği veya cevap vermek istemediği sorulara, yanlış veya eksik cevap verebilir. Siyaset ve din gibi toplumsal konularda herkes rahat olamayabilir. Verilen yanlış veya eksik cevaplar ise çalışmanın güvenilirliğini azaltacaktır.
3. *Gözlemci/Görüşmecilerden kaynaklanan hatalar:* Gözlemcinin, gözlem yaparken ilgisinin dağılması, başka konular ile ilgilenmesi, yorgunluk, uykusuzluk gibi etkenler yüzünden ortaya çıkabilecek hatalardır.

3.1.6. Veri Çeşitleri

Veriler, kullanım türlerine ve değişken sayılarına göre farklı alt birimlere ayrılırlar. Değişken sayısına göre tek değişkenli ve çok değişkenli veriler; ölçüm türlerine göre nicel ve nitel veriler; kayıt türüne göre yatay-kesit, zaman ve karma veriler olarak adlandırılırlar.

Değişken Sayısına Göre Veriler

1. **Tek değişkenli veriler:** Bu tip veri kümelerinde sadece tek bir değişken kullanılır. Tek değişkenli veriler için; bir sınıftaki öğrencilerin boy uzunlukları, Türkiye’deki şehirlerde bulunan okul sayıları örneklerini verebiliriz.
2. **Çok değişkenli veriler:** İki veya daha çok değişkenin kullanıldığı durumdur. Bir sınıftaki öğrencilerin boy uzunlukları, kiloları ve saç renkleri veya Türkiye’deki şehirlerde bulunan ilköğretim, lise, üniversite sayıları bu konu için birer örnektir.

Ölçüm Türlerine Göre Veriler

1. **Nicel veriler:** Sayılarla ifade edilebilen değişkenler nicel değişkenlerdir. Böyle bir değişken için toplanan veriler ise nicel verilerdir.
2. **Nitel veriler:** Sayılarla ifade edilemeyen değişkenler nitel değişkenlerdir. Böyle bir değişken için toplanan veriler ise nitel verilerdir.

Örnek:

Şehir	Hava Durumu	Sıcaklık	Nem	Basınç	Rüzgâr Yönü
Samsun	Yağışlı	15	70	35	Kuzey Batı
Zonguldak	Kapalı	18	64	33	Kuzey Batı
Sivas	Yağışlı	16	25	31	Güney
İzmir	Güneşli	25	56	31	Batı

Yukarıdaki örneğimizde; hava durumu ve rüzgâr yönü nitel veri, sıcaklık, nem ve basınç nicel verilerdir.

Kayıt Türüne Göre Veriler

1. **Yatay-Kesit veriler:** Zamanın belirli bir anında farklı birimlerden toplanan verilere yatay-kesit serisi verisi adı verilir. Yatay-kesit serisi verilerinde zaman sabittir fakat incelenen birimler veya değişkenler farklıdır. Veri değerlerinin kendi içinde anlamları olmasına karşın, verilerin sırasının bir an-

lamı yoktur. İllere göre belirli bir andaki kaza sayısı, ağaç sayısı veya günlük TL-Döviz kurları bu tip verilere örnek olarak verilebilir.

Örnek: Çeşitli döviz kurları

Döviz Cinsi	TL Karşılığı
1 ABD Doları	1,92
1 Avro	2,98
1 İngiliz Sterlini	3,32
1 Danimarka Kronu	0,26

2. **Zaman serisi verileri:** Değişkenlerin aldığı değerlerin gün, ay, yıl gibi belirli zaman birimlerine göre dağılım gösteren serilere zaman serisi denmektedir. Bir veya birden fazla değişkenin değerlerinin zamana göre değişimini belirten verilere ise zaman serisi verileri denmektedir. Zaman serisi verilerinde veri değerlerinin ve verilerin zaman içindeki sıralarının bir anlamı vardır. Aylık enflasyon verileri, yıllık ithalat-ihracat rakamları, mevsimlere göre işsizlik oranları veya yıllık doğum oranları, zaman serilerine örnek verilebilir.

Örnek: 2014 yılı aylık ABD Dolar kurları

Aylar	TL Karşılığı (Dolar)
Ocak	2,13
Şubat	2,26
Mart	2,21
Nisan	2,14
Mayıs	2,11
Haziran	2,09
Temmuz	2,09

3. **Karma veriler:** Zaman serisi ve yatay-kesit serilerinin birleşiminden oluşan verilere karma veri denir. Türkiye'nin bölgelere göre yıllık doğum oranları veya OECD ülkelerinin aylık petrol gelirleri örnek olarak gösterilebilir.

Karma veri, aynı yatay-kesit birimlerinin zamana göre değişimini gösteriyorsa bu tür veriler *panel veri* olarak adlandırılmaktadır. Yıllık doğum oranları Türkiye'deki bölgelere göre veriliyorsa, bu panel veridir.

Örnek: Ülkelere göre bilgisayar ve internet kullanıcı sayıları

	Bilgisayar Sayısı (Bin adet)		İnternet Kullanıcısı (Milyon kişi)	
	1990	2000	1990	2000
Almanya	3467	4129	39	47
Fransa	2438	3222	31	38
İngiltere	4781	5230	45	51
İspanya	3521	4300	27	33
Rusya	8779	10221	49	65
Portekiz	1923	3267	13	24

3.1.7. Verilerin Ölçme Düzeyleri

Ölçme; nesnelere, olaylara ve bireylere sayılar ve semboller verme süreci anlamını taşır. Bu süreçte, sayılar veya semboller arasındaki ilişkiler, ölçülen özellikler arasındaki ilişkileri yansıtacak şekilde olmalıdır. Fen bilimlerinde deney veya ölçüm yapma yoluyla veriler elde edilebilirken; sosyal bilimlerde ise yoruma dayalı olarak veriler elde edilebilir.

Ölçme türleri şunlardır:

1. Sınıflayıcı ölçme (Nominal)
2. Sıralayıcı ölçme (Ordinal)
3. Eşit aralıklı ölçme (İnterval)
4. Oranlı ölçme (Ratio)

1. Sınıflayıcı ölçme (Nominal)

Aynı özellikleri paylaşan insanları, olayları ya da diğer objeleri bir grup altında toplamak için kullanılır. Örneğin, insanları cinsiyetine, göz rengine ya da tuttıkları partiye göre sınıflandırmak. Ancak, mavi gözlü olanlar yeşil gözlü olanlardan iyidir ya da önce gelir denilemez. Araştırmacılar, pratikte grupları belirtmek için harf veya numaralar kullanabilir. Numara ve harfler bir kimlik olmaktan öteye gitmemektedir. Nesnelere verilen sayıların sayısal anlamları yoktur ve miktar belirtmezler. Gruplardaki elemanlar, frekans dağılımı olarak gösterilebilir.

2. Sıralayıcı ölçme (Ordinal)

Nesne ya da birimlerin herhangi bir özelliğe sahip oluş derecesine göre en yüksekte en düşüğe sıralanmasıdır. Böylece her bir değer diğerine eşit, diğ-

rinden küçük ya da diğerinden büyük olabilir. Örneğin, düşüncelerin en önemli olandan en önemsiz olana doğru sıralanması ya da testlerden alınan sonuçların derecelendirilmesi (Katılıyorum-Fikrim Yok-Katılmıyorum, Birinci-ikinci, çok iyi-iyi-kötü vs.). Bu ölçeklerde iki şey arasındaki fark eşit değildir. Mesela bir sınavdan alınan puanlara göre (20, 21 ve 80) sıralama yapıldığında ikinci ile üçüncü arasındaki fark ile birinci ile ikinci arasındaki fark eşit değildir. Başka bir deyişle, sıralananların birbirleri arasındaki farkları konusunda bir şey söylemek imkânsızdır. (100 metre finalinde ilk üçe giren yarışçıların sıralamasının bilinmesi fakat aralarında ne kadar süre farklarının olduğunun bilinmemesi gibi.)

3. Eşit aralıklı ölçme (Interval)

Bu tür ölçekler, birey ve durumlar arasındaki farkın miktarını göstermeye yöneliktir. Aralıklı ölçekler üzerinde hem ölçümün değeri, hem de ölçümler arasındaki farkın miktarı önemlidir. Aralıklar arasındaki sayısal değer, ölçeğin her tarafında aynıdır. Örneğin 5 ile 6 arasındaki farkla 18 ile 19 arasındaki fark eşittir. Bu tür ölçeklerde bir başlangıç noktası vardır, fakat bu başlangıç bağıl bir başlangıç noktasıdır ve ölçek eşit birimlere bölünmüştür. Termometreler, takvimler ve puanlar bu tür ölçeklere uygundur. Sıfır noktası sabit olmadığı için başlangıç noktası görecelidir. Burada sıfır noktası sayısallaştırmayı kolaylaştıran bir başlangıç noktasıdır, doğal ya da mutlak sıfır noktası değildir. Oda sıcaklığı 0 derecedir dendiğinde buradaki sıfır gerçek değildir, çünkü oda sıcaklığının olmadığı söylenemez.

4. Oranlı ölçme (Ratio)

Eşit oranlı ölçekler en yüksek düzeydeki ölçeklerdir. Daha önceki ölçeklerin tüm özelliklerini taşımakla beraber mutlak sıfır noktasına da sahiptir. Bu tür ölçekler aynı zamanda sıralı ve aralıklı ölçeklerdir. Birey ve durumlar arasındaki fark, oran biçiminde ifade edilir. Bir sayının iki katı, dörtte biri bu konu için örnek sayılabilir. Bu ölçeklerde mutlak bir sıfır noktası vardır, Yani ölçülen şey gerçekten yoktur. Metre, kilogram, saat gibi ölçümler. Ancak, eğitimde ölçümler oranla ifade edilmez. Örneğin bir öğrenci ya daha çok grupta çalışmayı sever ya da daha az. Bu iki kat daha çok seviyor ya da üç kat daha az seviyor diye ifade edilmez.

Ölçekler ise, deneklerin belli veriler temel alınarak sınıflandırıldıkları araç veya mekanizmalardır. Nesnelere sayılar vermede ve bu sayıların kullanılmasında uyulması gerekli kuralları ve kısıtlamaları belirtmek için kullanılır. Verilerimizin ölçme düzeyi; seçeceğimiz istatistiksel teknikler konusunda da bizlere yol gösterici olmaktadır.

Ölçek türleri ise şu şekildedir:

1. Sürekli ölçekler
2. Tekil ölçekler
3. Çoklu ölçekler

1. Sürekli ölçekler: Bu ölçeklerde cevaplayıcı, iki uç arasında uzanmakta olan bir çizgi üzerindeki uygun bir yere işaret koyar. Değerlendirme aşamasında araştırmacı cevap çizgisini gereken sayıda aralığa böler ve cevaplayan kişinin cevabını rakamsal olarak ortaya koyar. Oluşturulması kolay fakat skorların belirlenmesi zordur. Ayrıca çalışmanın güvenilirliği, ölçeği hazırlayana bağlı olması da bir soru işaretidir.

2. Tekil ölçekler: Tek madde veya sorudan oluşan bir ölçektir. Uygulamada beş farklı çeşidi vardır.

Kategorili ölçekler, cevap seçenekleri kategoriler halinde verilerek hazırlanır. Bu yöntemde, kategoriler arasında çakışma olmamalıdır. Her denek sadece bir seçeneğe ait olmalıdır.

Sıralama ölçeklerinde deneklere inceleme konusuyla ilgili özellikler listesi verilir ve onlardan listedeki özellikleri tercih, önem sırası veya başka bir ölçüte göre sıralaması istenir.

Sabit-toplam ölçekleri deneklerden, kendilerine verilen 100 tam puanın önem derecesi veya hoşlanma seviyesine göre listede verilen özelliklere dağıtılması istenir. Verilen puanların toplamı tam olarak 100 olmak durumundadır.

Şekilsel ölçekler özellikle okuma yazma bilmeyen ve küçük çocuklar üzerinde yapılır. Bu ölçek yönteminde cevaplar kelime halinde değil, şekiller ve oyunlarla verilmektedir.

İkili-karşılaştırmalı ölçeklerinde, deneklerden verilen ikililer arasında seçim yapması istenmektedir. Birbirlerine en yakın iki seçenek, en iyi cevabı bulana kadar sorulmaya devam edilir.

3. Çoklu ölçekler: Birden çok maddeden meydana gelen ve maddelerin aynı ölçeğe göre değerlendirildiği ölçekler, çoklu ölçekler olarak adlandırılır. Bu ölçekler, tek boyutlu olarak ölçülmesi mümkün olmayan çok boyutlu kavramların ölçülmesinde kullanılırlar.

Likert ölçeği, anket çalışmalarında en yaygın kullanılan ölçek türüdür. İnsanların belirli bir konuya ilişkin tavırlarını ölçmek için tasarlanmış bir yöntemdir. Anket yoluyla kişilere sorulan ve genelde 5'li ölçekten oluşan Likert ölçeği,

sosyal bilimlerde en sık kullanılan veri elde etme yöntemi olarak kullanılmaktadır. 5'li likert ölçeği ile ilgili şu örneği verebiliriz.

- ✓ Kesinlikle Katılıyorum
- ✓ Katılıyorum
- ✓ Kararsızım
- ✓ Katılmıyorum
- ✓ Kesinlikle Katılmıyorum

3'lü likert ölçeği örneği de şu şekildedir:

- ✓ Memnunum
- ✓ Kararsızım
- ✓ Memnun değilim

Sonuçları değerlendirirken, ifadeler arası ilişki incelenebilir ve skorların toplamı üzerinden değerlendirme yapılabilir. En önemli sakıncası anket doldurmanın süresinin uzun olmasıdır.

Semantik farklılıklar ölçeği; mamul, marka veya firma imaj çalışmalarında yaygın olarak kullanılan bir ölçek türüdür. İncelenen nesne/marka/ürüne ait çeşitli özelliklere ilişkin nitelendirmeler zıt kutuplara gelecek şekilde yedi noktalı bir doğru üzerine yerleştirilmiştir. Denek kendine uygun bir yere işaret koyar. Bu ölçek ile aralık seviyesinde ölçüm yapmak amaçlanmıştır. 1-7 şeklinde veya -3; +3 şeklinde örnekleri vardır.

3.2. Seriler

Derleme sonucunda elde edilen veriler, bir veri yığını oluşturur. Böyle bir yığından, belirlenen amaçlar doğrultusunda ihtiyaç duyulan bilgilerin elde edilebilmesi, ancak verilerin belirli esaslara göre düzenlenmesiyle mümkün olur.

Bunun için de ilk akla gelen, verileri büyüklüklerine göre sıralamaktır. Böyle bir sıralama sonucu elde edilen rakamlar dizisine, istatistik serisi adı verilir. Seriler gözlem sonuçlarını zaman ve mekân vasıfları ile maddi vasfın sıklığına göre sıralanmış sayı dizileridir.

Bir istatistik serisi, gözlem amaçlarına uygun değişken ya da değişkenlerin aldıkları değerlerden (sıklardan) oluştuğu için, ilgilenilen yığın olayın gerek yapısı, gerek değişimlerine ilişkin ayrıntılı ipuçları içerir.

1. Zaman Serileri

Eğer gözlem sonuçları yıl, ay, hafta, gün ya da saat gibi bir zaman değişkeninin şıklarına göre sıralanırsa, oluşturulan seriye *zaman serisi* adı verilir. Yıllara göre ülke nüfusları ve belirli bir noktada günün saatlerine göre trafik yoğunluğu, bu tür serilere örnek olarak gösterilebilir.

Örnek: 2013 Türkiye aylık en yüksek yağış miktarları (mm.) tabloda verilmiştir.

Aylar	Yağış Miktarı	Aylar	Yağış Miktarı
Ocak	142,6	Temmuz	164,8
Şubat	68,2	Ağustos	204,6
Mart	169	Eylül	80
Nisan	135	Ekim	153,5
Mayıs	94,8	Kasım	160,8
Haziran	75,5	Aralık	70

2. Mekân Serileri

Eğer gözlem sonuçları ülke, bölge, şehir ya da köy gibi bir mekân (yer) değişkeninin şıklarına göre sıralanırsa, elde edilen seriye *mekân serisi* adı verilir. Şehirlere göre elektrik tüketimi, bölgelere göre tahıl üretimi bu tür serilere örnek olarak gösterilebilir.

Örnek: Bazı illere ait rakım değerleri tablosu

İl	Rakım
Samsun	4
Zonguldak	10
Sivas	1285
İzmir	2
Ankara	890
Malatya	998
Erzurum	1853

3. Dağılma Serileri

Gözlem sonuçlarının maddesel bir değişkenin sıklıklarına göre sıralanmasıyla oluşturulan serilere, *dağılma serileri* adı verilir.

Örnek: İstatistik sınav sonucu oluşan not aralığına göre öğrenci sayıları

Not Aralığı	Öğrenci Sayısı
1-20	6
21-40	14
41-60	12
61-80	15
81-100	11

Dağılma serileri ana çizgileriyle *nicel* ve *nitel* dağılma serileri olmak üzere ikiye ayrılır. Nicel dağılma serileri de basit seriler (diziler), frekans serileri ve sınıflandırılmış (gruplandırılmış) seriler olmak üzere üç alt başlık altında incelenebilir.

4. Bileşik Seriler

Gözlem sonuçlarının iki veya daha fazla değişkene göre gösterildiği serilerdir.

Örnek: 2014 Dünya Basketbol Şampiyonası C grubunda atılan ve yenilen sayılara göre sonuç tablosu

Not Aralığı	Atılan Sayı	Yenilen Sayı
ABD	511	345
Türkiye	365	372
Dominik C.	347	386
Y. Zelanda	347	376
Ukrayna	344	369
Finlandiya	342	408

3.3. Sınıflama ve Gruplama

Toplanan verilerin çok sayıda olması veya karmaşık halde yer almalarından dolayı, verileri tasnif etmek, belirli bir düzen içinde analize hazırlamak gerekebilir. Bu gibi durumlarda verileri belirli bir sıraya ve düzene sokmak için sınıflama ve gruplama tekniklerinden yararlanılır.

3.3.1. Sınıflama

Araştırmada toplanan veriler genellikle düzensiz haldedirler. Verilerin işlenmesi denilince akla bunların sınıflandırılması veya gruplandırılması gelir. Sınıflama; incelenen olayın aynı şikkına sahip birimleri, kümeler halinde ayırmaktır.

Olayların çeşitli seçeneklerinin anakitlede kaç defa tekrarlandığını ifade eden sayılara *frekans* denir. Frekans f_i ile gösterilir. Frekansların kullanılmasıyla elde edilen serilere sınıflandırılmış seriler denir. Frekans toplamının, en başta verilen örnek hacmine eşit olduğuna dikkat etmemiz gerekir.

Örnek: *Ekonometri bölümü 1. sınıfta okuyan 28 öğrencinin yaşları sorulmuş ve alınan cevaplar aşağıda verilmiştir. Verilen cevapları frekanslı seri haline dönüştürünüz.*

$x_i = 19, 21, 22, 21, 19, 20, 22, 21, 20, 20, 20, 18, 19, 23, 24, 20, 21, 22, 22, 23, 25, 19, 19, 20, 24, 23, 22, 19$

Çözüm: Tablo, hangi değerden kaçar tane olduğu sayılarak ilgili değerın karşısına yazılarak hazırlanır.

x_i	f_i
18	1
19	6
20	6
21	4
22	5
23	3
24	2
25	1

Not: $\sum f_i = 28$ olduğuna dikkat edelim.

Örnek: Bir para 12 defa atılmış ve sonuçlar aşağıda verilmiştir. Verilen sonuçları frekanslı hale getiriniz.

x_i = Tura, Yazı, Yazı, Yazı, Tura, Yazı, Yazı, Yazı, Yazı, Tura, Yazı, Tura

Çözüm:

x_i	f_i
Yazı	8
Tura	4

Sınıflandırmada karşılaşılan sorunlar:

- Ele alınan olayın şıklarının çok sayıda olması. (Türkiye’de yaşayanların yaşları) (Burada gruptama yapmamız şart)
- Bir gözleme ait birden çok nitelik bakımından sınıflama yapma durumu. (Cinsiyet, medeni hal, yaş gibi birden çok nitelik)

3.3.2. Gruplama

Birbirlerine yakın değerleri gruplar halinde toplama işlemine *gruplama* denir. Meslek grupları, yaş kategorileri, boy uzunlukları gibi, bir kitlede çok fazla çeşit içeren veriler için uygundur. Buradaki en önemli konu; gruptamada kullanılan ~~sınıf aralıkları~~ neye göre belirlenmelidir, sorusudur. Kesin kurallar belirlememekle beraber şu kurallar tavsiye edilir.

- Genellikle 5 ile 20 arasında olması tavsiye edilir.
- En alt ve en üst değerdeki veriler, dışarıda kalmamalıdır.
- Aralıklar eşit olmalıdır. Aralıkların eşit olmadığı durumlarda, genellikle son sınıfın aralık değerinin farklı alınması uygulamada sıkça kullanılmaktadır.

Verileri gruptamada yararlanabileceğimiz uygun bir aralık katsayısı saptamak için kesin bir kural olmamasına rağmen aşağıda verilen formül yardımıyla grup aralıkları hesaplanabilir.

Dağılım Genişliği (Ranj) = En büyük değer – En küçük değer

$$R(Ranj) = X_{Maks} - X_{Min}$$

$$\text{Sınıf Sayısı } (k) = 1 + (3,3) \cdot \log n$$

n = örnek hacmi

$$\text{Sınıf aralığı } (c) = \frac{R}{k}$$

$$\text{Tahmini Aralık Sayısı} = \frac{\text{Ranj}}{\text{SINIF SAYISI}}$$

Örnek: Verilen basit seriyi gruptandırınız.

$x_i = 12, 14, 15, 15, 16, 18, 18, 21, 22, 23, 25, 25, 25, 26, 30, 32, 33, 35, 36, 40, 42$

Çözüm:

$$R(\text{Ranj}) = X_{\text{Maks}} - X_{\text{Min}} = 42 - 12 = 30$$

$$n = \text{örnek hacmi} = 21$$

Sınıf sayısı $(k) = 1 + 3,3 \cdot \log(n) = 1 + 3,3 \log 21 = 1 + 3,3 \cdot 1,322 = 5,3626$ olarak hesaplanır. Kendisine en yakın tamsayı değerini aldığımızda, istenen sınıf sayısı 5 olarak bulunur. Sınıf sayısını bulduktan sonra her bir grup için tahmini aralık sayısı hesaplanır.

$$\text{Tahmini Aralık Sayısı} = \frac{\text{Ranj}}{\text{SINIF SAYISI}} = \frac{30}{5} = 6$$

Gruplama yaparken, hiçbir seri elemanı dışarıda bırakılmamalıdır. Bu örneğimizde alt sınır olan 12 değerine 6 ekleyerek birinci sınıfı oluştururuz. Aynı işlemi, her sınıfın son değerine 6 ekleyerek kalan sınıfları bulabiliriz. Toplamda 5 adet sınıf bu şekilde oluşturularak, gruptandırma işlemi tamamlanır. Genel uygulama olarak alt sınır değerine eşitlik verilir (kesin bir kural olmamakla beraber) ve diğer gruplara da aynı işlem uygulanır. Sınıf aralığı belirlendikten sonra, o aralıkta yer alan değerleri sayarak frekans kısmına yazarız. (Seriye ait en büyük değer son sınıfın üst değerine eşitse, üst değere de eşitlik işareti konulmalıdır.)

Sınıf Aralığı	Frekans
$12 \leq x < 18$	5
$18 \leq x < 24$	5
$24 \leq x < 30$	4
$30 \leq x < 36$	4
$36 \leq x \leq 42$	3

Frekans toplamının, örnek hacmine eşit olduğu unutmamak gerekir.

Örnek: Verilen basit seriyi gruplandırınız.

$$x_i = 124, 134, 132, 119, 133, 127, 136, 150, 112, 122, 134, 133$$

Çözüm:

$$R = 150 - 112 = 38$$

$$n = 12$$

$$k = 1 + 3,3 \cdot \log_{10} 12 = 4,56 \approx 5$$

$$\text{Tahmini Aralık Sayısı} = \frac{\text{Ranj}}{\text{SINIF SAYISI}} = \frac{38}{5} = 7,6 \cong 8 \text{ olarak hesaplanır.}$$

Sınıf Aralığı	Frekans
$112 \leq x < 120$	2
$120 \leq x < 128$	3
$128 \leq x < 136$	5
$136 \leq x < 144$	1
$144 \leq x \leq 152$	1

Gruplandırma yaparken yukarıda verdiğimiz formül ile sınıf aralığı ve sayısı bulunacak diye bir kural olmadığından bahsetmiştik. Araştırmacının keyfi kararları ile de gruplandırma yapılabilir. Araştırmacı grup sayısını ve sınıf aralığını belirterek, hiçbir gözlem dışarıda kalmamak kaydıyla basit seriyi gruplandırılmış seri haline dönüştürebilir.

Örnek: Verilen basit seriyi 4 sınıf ve her sınıf aralığı 13 olacak şekilde gruplandırınız.

$$x_i = 45, 46, 38, 67, 54, 67, 78, 65, 65, 71, 44, 56, 60, 64, 57, 48, 53, 67, 88, 74, 39, 51, 65, 59, 44, 65, 77, 83, 87, 39$$

Çözüm: Toplamda 30 gözlem vardır. Bu gözlemlerin en büyüğü 88, en küçüğü ise 38'dir. Verileri küçükten büyüğe doğru sıralayarak, yapacağımız işlemler için kolaylık sağlayabiliriz.

$$x_i = 38, 39, 39, 44, 44, 45, 46, 48, 51, 53, 54, 56, 57, 59, 60, 64, 65, 65, 65, 65, 67, 67, 67, 71, 74, 77, 78, 83, 87, 88$$

Sınıf Aralığı	Frekans
$38 \leq x < 51$	8
$51 \leq x < 64$	7
$64 \leq x < 77$	10
$77 \leq x < 90$	5

Örnek: Verilen basit seriyi gruplandırınız.

$$x_i = 23, 34, 45, 45, 56, 36, 66, 88, 101, 26, 40$$

Çözüm: Önsel bir bilgi verilmediği için, daha önce verilen formül yardımıyla gruplandırma yapabiliriz.

$$R(Ranj) = 101 - 23 = 78$$

$$n = 11$$

$$k = 1 + 3,3 \log 11 = 4,43 \cong 4$$

$$\text{Tahmini Aralık Sayısı} = \frac{Ranj}{SINIF SAYISI} = \frac{78}{4} = 19,5 \cong 20 \text{ olarak hesap-}$$

lanır.

Sınıf Aralığı	Frekans
$23 \leq x < 43$	5
$43 \leq x < 63$	3
$63 \leq x < 83$	1
$83 \leq x < 103$	2

3.3.2.1. Kontenjans Tabloları

Verileri iki veya daha fazla nitel değişkene göre sınıflandırdığımızda, oluşacak sonuçları görmemize yarayan tablolara kontenjans tabloları adı verilir. Çapraz tablo olarak da adlandırılan kontenjans tablosu, genellikle 2 değişkenin birlikte etkisini görmek için kullanılır. Kontenjans tabloları, her 2 değişkenin göre yüzde değişimlerini görmek açısından kolaylık sağlamaktadır.