

İSTATİSTİĞE GİRİŞ

İstatistik, *deskriptif (tasviri)* ve *indaktif (tahlili)* istatistik olmak üzere ikiye ayrılmaktadır.

Deskriptif (Tasviri) İstatistik Tasviri: Herhangi bir konuyla ilgili verilerin toplanması, düzenlenmesi, özetlenmesi ve söz konusu verilerin tablo ve grafikler hâlinde gösterilmesi ile ilgilenir. Frekans dağılımları, merkezî eğilim ölçüleri (aritmetik ortalama, mod, medyan, ...), dağılım ölçüleri (standart sapma, varyans, değişim aralığı, ...), çarpıklık ve basıklık ölçüleri gibi.

İndaktif (Tahlili) İstatistik: İlgilenilen konuyla ilgili tüm veriler arasından seçilen alt veriler kullanılarak analizlerin yapılması ve bu analizler ile elde edilen sonuçlar kullanılarak tüm birimler hakkında yorum yapılması ve bir karara bağlanması ile ilgilenir. Bu tanımdan yola çıkarak indaktif istatistik, tahlili istatistik olarak da adlandırılmaktadır. Örneklem teorisi, hipotez testleri, regresyon ve korelasyon analizleri gibi.

Anakütle: Bir istatistiki araştırmada, araştırmaya konu olan bütün birimlere anakütle denir

Sınırlı anakütle, İktisadi ve İdari Bilimler Fakültesi'nin birinci sınıfında okuyan öğrencilerinin sınav notları ilgili araştırma örneğinde olduğu gibi bu öğrencilerin sayısı tespit edilebildiği anakütleleri ifade etmektedir. Kısacası araştırma konusu ile ilgili birimlerin çerçevesi çizilebiliyorsa bu anakütle **sınırlı anakütledir**. Sınırsız anakütle kavramı ise, Karadeniz'deki hamsiler ile ilgili yapılan araştırma örneğindeki gibi hamsilerin tamamının sayısı tespit edilemediği durumlarda karşımıza çıkmaktadır. Araştırma konusu ile ilgili birimlerin çerçevesi çizilemediği durumlar **sınırsız anakütleyi** ifade etmektedir. Karadeniz'deki hamsilerin sayısı sınırsız anakütle için örnek teşkil ettiği gibi, bir ülkede roman okuyan şahısların sayısı, bir ildeki herhangi bir yemeği seven insanların sayısı, bir ülkede seyahat etmeyi sevenlerin sayısı gibi konularda anakütleyi oluşturan birimlerin sayısı tespit edilemediği için sınırsız anakütle için örnek teşkil etmektedir.

Örneklem ve Örnek: Tüm anakütleye ulaşmak mümkün olmayabilir. Bu gibi durumlarda anakütleden tesadüfi yöntemlerle anakütle birim sayısından daha az sayıda birimlerin seçilme işlemine **örneklem** denir. Anakütleden örneklem yardımıyla seçilen birimler ise **örnek** olarak ifade edilmektedir.

Anakütle ve Örnek Hacmi: **Anakütle hacmi,** anakütleyi oluşturan birimler topluluğudur ve genellikle N ile gösterilir. **Örnek hacmi** ise örneğe seçilen birim sayısıdır ve n ile gösterilir. Eğer araştırma İktisadi ve İdari Bilimler Fakültesi için yapılıyorsa, fakültede okuyan 8000 öğrenci anakütle hacmini, bu öğrenciler içerisinde tesadüfi olarak seçilen 100 kişilik öğrenci grubu ise örnek hacmini ifade etmektedir. Bu tip bir örneğe **şans örneği** denir. Örneklem yapmanın temel amacı anakütleden seçilen örnekler yardımıyla anakütle hakkında bilgi elde etmektir. Örnekten elde edilen veriler kullanılarak elde edilen sonuçlar bütün anakütleye genelleştirilebilir. Anakütleden tesadüfi yöntemlerle anakütle birim sayısından daha az sayıda seçilen örnek yardımıyla tahminler yapılarak anakütle hakkında karar verilir

2. Ölçüm Türüne Göre Veriler: Veriler, ölçüm türlerine göre nitel veriler (kalitatif) sayısal olmayan veriler nicel veriler (kantitatif) sayısal olan veriler şeklinde genel olarak iki şekilde sınıflandırılır.

2.1. Nitel veriler: Değişkenin vasfı ile ilgili sayısal olmayan bilgilerdir. Bir birimin çeşitli kategorilerden hangisinde yer aldığını basit bir şekilde tespit edebildiğimiz veri türüdür. Nitel veriler sınıflayıcı (nominal) ve sıralayıcı (ordinal) olmak üzere iki şekilde incelenebilir:

2.1.1.Sınıflayıcı (Nominal) Veriler: Bu tür veriler için elde edilen değerler sayısal bir büyüklük ifade etmezler. Örneğin; araştırma konusuna göre sorulan cinsiyet, medeni durum, saç rengi, göz rengi, mesleği, okuduğu bölüm, doğduğu il, yaşadığı coğrafi bölge gibi değişkenlere verilen cevaplar nominal verilerdir.

2.1.2.Sıralayıcı (Ordinal) Veriler: Sıralayıcı verilerde ilgili değişkenin aldığı değerler açısından birbirlerine üstünlükler ya da önem derecesine göre bir sıralama söz konusudur. Bir yarışa katılan koşucular yarışı bitirme zamanlarına göre, birinciden sonuncuya kadar sıralanırken; belirli bir mamül hakkında tüketicilerin beğenileri, "Hiç beğenmedim, az beğendim, kararsızım, beğendim, çok beğendim." şeklinde sıralanabilir. Ordinal veriler yalnızca ">" ve "işlemleri için sayı gibi değerlendirilebilir, bunlar dışında matematik işlemler uygulanamaz.

Söz konusu gözlemler mekânâ göre sınıflandırılarak elde ediliyorsa bu serilere **mekân serileri** denir. Mekân (yer) serisi, araştırma konusuna ait gözlem sonuçlarını il, ilçe, bölge, ülke, kıta gibi yer itibarıyla gösteren serilerdir. İllere göre nüfus sayıları, bölgelere göre doğalgaz tüketim miktarları, farklı bölgelerdeki üniversitelerdeki öğrenci sayıları, çeşitli üniversitelerdeki bilimsel yayın sayıları mekân serilerine örnek olarak verilebilir.

2.1.3. Gruplandırılmış seriler Verilerin gruplandırılması, özellikle denek sayısı fazla olduğunda, hem veriler üzerinde yapılacak hesaplamaları kolaylaştırır hem de verilerin kolay anlaşılır biçimde okuyucuya sunulmasını sağlar.

Gruplandırılmış bir seri altı adımda oluşturulabilir:

Adım 1: Elde edilen veriler basit seri hâline getirilir.

Adım 2: Verilerin en büyüğünden en küçüğü çıkarılarak değişim aralığı bulunur. Değişim Aralığı (K) = $X_{max} - X_{min}$ Adım 3: Olayın kaç grupta inceleneceği (m) belirlenir.

Adım 4: Sınıf büyüklüğü (A) belirlenir. Sınıf büyüklüğü dağılımdaki en büyük değerden en küçük değer çıkarılıp grup sayısına bölünmesiyle elde edilir. Sınıf büyüklüğü aynı zamanda sınıf üst sınırı ile alt sınırı arasında farktır. $A = k/m$ Adım 5: Serideki en küçük değere sınıf büyüklüğü eklenerek 5. adımda serideki en büyük değere ulaşacak şekilde gruplar oluşturulur.

Adım 6: Her guruba düşen değer sayısı belirlenir

İki tip gruplandırılmış seri hazırlanabilir: *Sürekli gruplandırılmış seriler ve kesikli gruplandırılmış seriler*. Yukarıda gruplandırılmış seriye **sürekli gruplandırılmış seri** denir. Çünkü bir sınıfın bittiği noktada bir başka sınıf başlamaktadır. Sınıflar arasında boşluk bulunmamaktadır. Ancak bir sınıfın bittiği nokta ile bu sınıfın peşi sıra gelen sınıfın başlama noktası arasında boşluk varsa bu tür gruplandırılmış seriye **kesikli gruplandırılmış seri** denir.

Gruplandırılmış serilerde her sınıfın alt ve üst değerlerine **sınıf uçları** denir

Kesikli gruplandırılmış serideki sınıf uçları ise; 5-10, 11-16, 17-22, 23-28 değerleridir. Bu ikili rakamlardan birinciler alt uçlar, ikinciler ise üst uçlardır.

Kesikli gruplandırılmış serilerde bir grubun üst ucu ile bir sonraki sınıfın alt ucu toplanıp ikiye bölünerek o sınıfın üst sınırı bulunmuş olur. Sürekli gruplandırılmış seride sınıf uçları aynı zamanda sınıf sınırlarıdır.

Her sınıfın alt ve üst sınıf sınırları veya uçları toplanarak ikiye bölündüğünde sınıfa ait sınıf değeri elde edilmiş olur. Sınıf değeri, ilgili sınıfı temsil eden değerdir.

Grafik, “gözlem değerlerinin matematiksel ve bilimsel temellere sahip şekiller hâlinde ifade edilmesi” şeklinde tanımlanabilir. Grafikler gözlem değerlerinin daha kolay anlaşılmasını sağlamaktadır.

Histogram, gruplandırılmış seriler için oluşturulan bir dikdörtgenler dizisidir. **Histogram**; yatay eksen de değişkenin aldığı değerlerin, dikey eksen de ise frekansların bulunduğu ve her aralığın frekansı ile orantılı yüksekliklere sahip dikdörtgenlerin gösterildiği bir yoğunluk grafiğidir. Histogram çizilirken yatay eksen de gruplandırılmış serinin sınıf sınırları, dikey eksen de ise frekanslar yer almaktadır.

İlk kez İtalyan Ekonomist Pareto (1848-1923) tarafından bulunan Pareto grafiği, söz konusu bilim adamının adıyla anılmaktadır. Pareto grafiği gelir dağılımları ile ilgili çalışmalarda tespit edilmiştir. Grafik hata ve maliyet analizleri için kullanılan basit bir yöntemdir. **Pareto grafiği**; bozuk ürünler, tamirler, arızalar, talepler, noksanlıklar veya kazalar ile mali kayıplar ve bunların sebepleri gibi olayların görsel olarak meydana gelme frekanslarını gösteren bir tür frekans dağılım grafiğidir. Pareto grafiği bir sorunu oluşturan nedenleri önem sırasına göre sıralayarak, önemlileri önemsizlerden ayırt etmeye ve dikkatleri önemli nedenler üzerinde toplamaya imkân vermektedir. Grafikte soldan sağa doğru gidildikçe küçülen sütunlara sahip veri sınıfları sıralanmaktadır. Pareto grafiğinde en önemli problemler sol tarafta yer alırken, önemsiz olanlar sağ tarafta yer alır. Pareto grafiği bazı özellikleri bakımından histograma benzemektedir. Histogramdan şu özelliği ile ayırt edilebilir: Pareto grafiğinde yatay eksen kategorik verileri gösterirken, histogramda yatay eksen numerik verileri göstermektedir. Bazı durumlarda pareto grafiğinde kümülatif eğri gösterilmektedir. Söz konusu eğri soldan sağa doğru eklenen veri sınıflarının toplamını göstermektedir.

Sınav öncesinde son tekrar için hazırladığımız bu Gümüş Not;

Altın Notlar ve Altın Sorular'dan hazırlık yapan öğrencilerimizin anlayacağı şekilde kısaltılarak hazırlanmıştır.

www.akademiaof.com

Aritmetik ortalama: Serideki tüm değerlerin toplam eleman sayısına bölünmesi ile hesaplanır. Frekans ve sınıflı serilerde aritmetik ortalama ise serinin toplam değerinin toplam frekansa (birim sayısına) bölünmesi ile bulunur. Ortalama seride değişkeni ifade eden harfin üzerine çizgi çizilerek gösterilir. Aritmetik ortalama serideki bütün rakamlardan etkilenen bir ortalamadır. Gruplandırılmış serilerde X değerleri sınıf orta noktalarını yani sınıf değerlerini gösterirler. X değerlerine ait tartılar varsa frekanslar yerine tartılar kullanılarak tartılı aritmetik ortalama hesaplanır.

Geometrik ortalama (G), serideki n tane birimin çarpımının n' inci dereceden kökü alınmak suretiyle hesaplanır. Seride sıfır veya negatif değer varsa geometrik ortalama hesaplanamaz. Geometrik ortalama, geometrik dizi şeklinde artış gösteren serileri en iyi temsil eden parametrik merkezî eğilim ölçüsüdür. **Geometrik dizi** bir sayının katlanarak değerler alması durumunda oluşan seridir.

Harmonik ortalama serideki birimlerin çarpıma göre terslerinin aritmetik ortalamasının tersidir. Seride sıfır veya negatif birim bulunması hâlinde harmonik ortalama kullanılmaz. Harmonik ortalama parametrik merkezî eğilim ölçülerinin en küçüğü olarak elde edilir.

Kareli ortalama, serideki birimlerin karelerinin aritmetik ortalamasının kareköküdür. Basit serilerde kareli ortalama hesaplanırken önce serideki değerlerin kareleri hesaplanır. Sonra kareli değerleri toplamı elde edilir. Daha sonra kareli değerlerin toplamı serideki değer sayısına bölündüğünde elde edilen değer karekökü alınarak kareli ortalama elde edilir. •Kareli ortalama parametrik merkezî eğilim ölçülerinin en büyüğü olarak elde edilir. Parametrik merkezî eğilim ölçüleri arasında $H < G < X < K$ şeklinde büyüklük ilişkisi vardır

Mod, incelenen bir seride en fazla tekrar eden ya da başka bir ifadeyle frekansı en yüksek olan gözlem değeridir. Bir seride birden fazla mod olabileceği gibi seride mod bulunmayabilir. Mod, grafiksel olarak gösterildiğinde grafiğin tepe noktasında olduğundan, **tepe değer** olarak da ifade edilebilir. Mod, grafikte de tespit edilebilir. Verilmiş olan grubun, histogramı üzerindeki frekans eksenini en büyük olan kutucuğun, üst köşe noktaları ile komşu kutucuklara değdiği en üst noktaların çapraz bir şekilde birleştirilmesi ile oluşan kesişim noktasının yatay eksenindeki karşılığı serinin modudur.

Veriler büyükten küçüğe veya küçükten büyüğe doğru sıralandığında serinin tam ortasına karşılık gelen değere **medyan ya da ortanca** denir. Büyüklük sırasına göre sıralanmış basit bir serideki veri sayısı n olmak üzere, n tek ise medyan $(n+1)/2$ ' inci değerdir. Örnek hacmi n çift ise medyan $n/2$ ' inci değer ile $(n/2)+1$ ' inci değer aritmetik ortalamasıdır. Sınıflandırılmış seriden medyan bulunabilmesi için öncelikle ...den az kümülatif frekans değerlerinin bulunması gerekir. Medyanı gösteren $n/2$ 'inci değeri ilk kez içerisinde bulunduran kümülatif frekansa sahip olan değer medyan değeridir.

Kantiller, bir seriyi 4, 10 ve 100 eşit parçaya ayırarak bu serideki değerlerin dörtte, onda ve yüzde ne kadarının belirli bir değere göre yerini saptamak için kullanılır.

Kartiller, büyüklük sırasına dizilmiş bir serinin dört eşit parçaya bölünmesi sonucu üç kartil bulunur. Küçükten büyüğe doğru sıralanan seriyi dört parçaya bölebilmek için üç bölen gerekir. Birinci kartil Q1, ikinci kartil Q2 ve üçüncü kartil Q3 ile gösterilir. Her bir kartil aralığı yaklaşık serideki rakamların %25'ini kapsar. Bir serinin ikinci kartili, medyandır. Basit serilerde; $Q1 = (N+1)/4$ ' üncü değerdir. $Q2 = 2(N+1)/4$ ' üncü değerdir. $Q3 = 3(N+1)/4$ ' üncü değerdir.

Seride bulunan gözlem değerlerinin aritmetik ortalamadan farklarının kareleri alınarak toplanırsa **en küçük fark kare toplamı** elde edilir. En küçük fark kare toplamı serbestlik derecesine bölünerek varyans değeri hesaplanır. Serbestlik derecesi toplam gözlem sayısının bir eksisidir. Örnek değerleri kullanılarak hesaplanan varyans değeri s^2 ile ifade edilirken, anakütledeki tüm değerler kullanılarak hesaplanan varyans değeri ise σ^2 ile ifade edilir

Kareli ortalamanın karesinden aritmetik ortalamanın karesi çıkarıldığında varyans değeri elde edilmektedir.