

Giriş

İstatistik Nedir?

İstatistik, verilerin toplanması, düzenlenmesi, özetlenmesi, analiz edilmesi ve yorumlanması ile uğraşan bir bilim dalıdır. Günümüzde bilgiye dayalı karar alma süreçlerinde istatistik vazgeçilmez bir araç haline gelmiştir. İster bireysel kararlar isterse büyük organizasyonların stratejik kararları olsun, çoğu zaman güvenilir veriler ve bu verilerin doğru analizi gereklidir.

İstatistik bilimi aşağıdaki işlevleri içinde barındırır:

- Verileri tanımlama ve özetleme: Büyük veri yığınlarının sistematik bir şekilde özetlenmesi, betimlenmesi ve anlam çıkarılması istatistiğin temel uğraşlarından biridir. Betimsel istatistik bu amaca uygun sayısal ve görsel yöntemler geliştirmiştir.
- Belirsizliğin ele alınması: İstatistik belirsizliğin bilimi olarak düşünülebilir. Olasılık teorisi, istatistiksel çıkarımın temelini oluşturur ve belirsizliğin incelenmesi için gereken bilimsel dili ve araçları sunar.
- Anakütleyle ilişkin çıkarım ve genellemeler yapma: Örneğin, yeni bir ürün geliştiren bir firma tüm tüketicilerden oluşan anakütlenin bu ürüne olan talebi ile ilgilenebilir. Ancak çoğu durumda anakütle yerine onu temsil eden bir örneklem üzerinden çıkarım ve genellemeler yapılır. Daha küçük bir tüketici alt kümesinden hareketle tüm pazara ilişkin genellemeler yapılabilir ve ürüne ilişkin bir pazarlama stratejisi geliştirilebilir.
- İlişkilerin analizi: İstatistik, değişkenler arasındaki ilişkilerin analiz edilmesi için çeşitli yöntemler içerir. Korelasyon ve regresyon analizi, iki değişken arasındaki ilişkinin doğasını anlamaya yardımcı olur. Örneğin, reklam harcamaları ile satışlar arasındaki ilişki bu yöntemlerle ortaya konabilir.
- Nedensellik analizi: İstatistik aynı zamanda nedensel ilişkilerin tahmin edilmesi için deneysel yaklaşımlar sunar. Deney tasarımı ve deneysel verilerin toplanması, belirli bir müdahalenin (tedavinin) hedef değişkenler üzerindeki etkisini analiz etmeye olanak tanır. Bu deneysel yöntemler, gözlemsel çalışmalardan farklı olarak, kontrol grupları ve rastgele atama gibi tekniklerle bir nedensellik ilişkisi kurmaya yardımcı olur.
- Geleceğe yönelik tahminler yapma (kestirim/öngörü): İstatistik, geçmiş verilere dayanarak gelecekteki olaylar hakkında tahmin yapmak için yöntemler geliştirir. Örneğin, bir şirket geçmiş yıllardaki satış verilerine dayanarak önümüzdeki yıl için satış tahminlerinde bulunabilir. Zaman serisi analizi, ekonometrik modeller ve makine öğrenimi yöntemleri bu süreçte kullanılabilir. İstatistik tüm bu yöntemlerin temelini oluşturur.
- Belirsizlik altında karar verme: Karar verme süreçlerinde belirsizlik her zaman mevcuttur. İstatistik, bu belirsizlikleri hesaba katarak olası sonuçların risklerini değerlendirir ve daha doğru kararlar alınmasına yardımcı olur. Örneğin, bir yatırım şirketi, borsadaki

dalgalanmaları göz önünde bulundurarak hangi hisse senedine yatırım yapacağına karar verirken istatistiksel analizlerden yararlanabilir.

İstatistiğin kullanım alanları

İstatistik hemen her alanda kullanılır. Örneğin:

- **Ekonomi:** Enflasyon, işsizlik oranları, büyüme oranları gibi ekonomik göstergelere ilişkin veri toplanması, bunların hesaplanması ve analiz edilmesi istatistiğin konuları arasındadır. Ayrıca, ekonometrik modellerin oluşturulması, tahmini ve yorumlanması için istatistiksel araçların kullanılması zorunludur. Örneğin, bir ülkenin GSYİH büyüme oranlarını veya gelir dağılımındaki değişimleri incelemek ve nedensel ilişkiler kurmak ekonometrik analizler gerektirir.
- **İşletme:** İş dünyasında istatistik, karar verme süreçlerini destekleyen önemli bir araçtır. Pazarlama stratejilerinin belirlenmesi, müşteri memnuniyeti anketlerinin analizi ve finansal risklerin yönetimi istatistiksel yöntemlerle gerçekleştirilir. Örneğin, işletmeler pazar araştırmaları ve müşteri verilerini kullanarak ürün geliştirme ve pazarlama stratejilerini şekillendirirler. Finans alanında, portföy yönetimi, risk analizi ve stok tahminleri gibi süreçlerde de istatistiksel yöntemler kullanılmaktadır. Ayrıca, işletme operasyonlarının etkinliğini ölçmek ve iyileştirmek için veri analitiği ve ileri istatistiksel modeller kullanılarak daha fazla tercih edilmektedir.
- **Sağlık Bilimleri:** Klinik çalışmalar, ilaç etkinliği testleri ve epidemiyolojik araştırmalar gibi alanlarda istatistiksel analizler büyük bir öneme sahiptir. İlaçların güvenliği ve etkinliği üzerine yapılan araştırmalar, rastgele kontrollü deneyler (RCT'ler) ile yürütülür ve sonuçların analizinde istatistiksel yöntemler kullanılır. Ayrıca, bulaşıcı hastalıkların yayılımını inceleyen epidemiyolojik çalışmalar, salgın hastalıkların kontrolü ve önlenmesi için gerekli olan verileri sağlar.
- **Sosyal Bilimler:** Toplumun genel eğilimlerini anlamak, kamuoyu yoklamaları ve anket analizleri istatistiğin önemli bir parçasıdır. Örneğin, seçim anketleri veya tüketici davranışlarıyla ilgili araştırmalar, istatistiksel analizlerle desteklenir. Sosyal bilimlerde araştırmalarında nedensel ilişkilerin belirlenmesi de istatistiksel modellerle yapılır, bu da kamu politikalarının ve sosyal programların etkisini anlamada önemli bir rol oynar.
- **Mühendislik:** Mühendislikte istatistik, üretim süreçlerinde kalite kontrol ve güvenilirlik analizleri için vazgeçilmezdir. Üretim hatlarının verimliliğini artırmak, arızaları önlemek ve ürün kalitesini sağlamak için mühendisler, istatistiksel süreç kontrolü ve güvenilirlik mühendisliği gibi çeşitli yöntemler kullanırlar.

İstatistiksel süreç ve bilimsel yöntem

İstatistiksel analiz genellikle şu adımları içerir:

- Veri Toplama: Doğru ve güvenilir veri toplama süreci, istatistiksel analizlerin temelini oluşturur. Anketler, deneyler veya gözlemlerle veri toplanabilir.
- Veri Düzenleme: Toplanan veriler analiz edilebilir hale getirilir. Eksik veriler tamamlanır, hatalı veriler düzeltilir.
- Veri Analizi: İstatistiksel teknikler kullanılarak veriler incelenir. Betimsel ve çıkarımsal istatistiklerle veri hakkında sonuçlar çıkarılır.
- Sonuçların Yorumlanması: Elde edilen bulguların anlamı değerlendirilir ve karar alma süreçlerinde kullanılır.

Bilimsel yöntem, gözlemler ve deneylerle hipotezleri test etmeyi içerir. İstatistik, bu süreçte kullanılan temel yöntemleri içerir ve yeni yöntemler geliştirir. Bilimsel araştırma süreçleri genel olarak şu aşamalardan oluşur:

- Hipotez Oluşturma: Bilimsel araştırmaların ilk adımı araştırma sorusunu ve hipotezlerini ortaya koymaktır. Araştırmacı teorik modellere, mevcut bulgu ve gözlemlere dayanarak bir hipotez ileri sürer.
- Veri Toplama: Hipotezi test etmek için deneyler veya gözlemler yoluyla veri toplanır.
- Veri Analizi: Toplanan veriler istatistiksel analizlerle değerlendirilir.
- Hipotezin Test Edilmesi: Veriler ışığında hipotez kabul veya reddedilir.

Deneysel ve gözlemsel araştırmalar bu sürecin iki temel bileşenidir:

- Deneysel Araştırmalar: Araştırmacının müdahalesiyle yürütülen kontrollü deneyleri kapsar.
- Gözlemsel Araştırmalar: Doğal ortamda gözlemlerle veri toplanan araştırmaları içerir.

İstatistik, Veri Bilimi ve Yapay Zeka

İstatistik, günümüzde hızla gelişen veri bilimi, makine öğrenimi, veri madenciliği ve yapay zeka (AI) gibi interdisipliner alanların temel taşlarından biridir. Bu alanlar, büyük veri kümeleri ile çalışmak ve bu verilerden anlamlı sonuçlar çıkarmak için istatistiksel yöntemlerden geniş ölçüde yararlanır.

Veri Bilimi (Data Science): Veri bilimi, büyük veri kümelerinden bilgi çıkarma amacını taşıyan ve istatistik, matematik, bilgisayar bilimi ve diğer bilim dallarının kesişiminde çok disiplinli bir çalışma alanıdır. İstatistik, veri bilimcilerin veri temizleme, modelleme, analiz ve görselleştirme

süreçlerinde kullandıkları temel araçları içerir. Ayrıca, istatistiksel çıkarımlar, veri biliminin hipotez testi ve güven aralıkları gibi analiz süreçlerine rehberlik eder.

Makine Öğrenimi (Machine Learning): İstatistik, makine öğrenimi ve istatistiksel öğrenme algoritmalarının geliştirilmesinde büyük bir rol oynar. Regresyon, sınıflandırma, kümeleme gibi temel makine öğrenimi algoritmaları istatistiksel yöntemlerle yakından ilişkilidir. İstatistik, ayrıca model performansını değerlendirme ve sonuçların genellenebilirliğini sağlama süreçlerinde de kullanılır.

Veri Madenciliği (Data Mining): Veri madenciliği, büyük veri kümelerinden anlamlı desenler ve ilişkiler keşfetme sürecidir. Bu alanda kullanılan pek çok yöntem, istatistiksel teknikler üzerine kuruludur. Örneğin, kümeleme algoritmaları, ilişki kurma ve hipotez oluşturma gibi işlemler, istatistiksel modellere dayanır.

Yapay Zeka (Artificial Intelligence - AI): Yapay zeka, insan benzeri karar verme süreçlerini benzeten (simülasyonunu yapan) sistemler geliştirmeyi hedefler. Bu sistemlerin temelinde, genellikle makine öğrenmesi, istatistiksel öğrenme ve olasılık teorisi yatar. AI uygulamalarında belirsizlikleri yönetme, riskleri hesaplama ve sonuçları tahmin etme gibi işlemler için istatistiksel yöntemler hayati öneme sahiptir. Örneğin, doğal dil işleme (NLP) alanında büyük metinlerde kelime sıklığı ve sentiment analizinde, hangi kelimelerin daha anlamlı olduğunu, ve olumlu veya olumsuz duygudurumu belirlemek için istatistiksel yöntemler kullanılır.

Sonuç olarak, bu alanlar istatistikten doğrudan beslenmekte ve her biri verinin gücünden yararlanarak daha iyi kararlar almak ve tahminlerde bulunmak için istatistiksel araçları kullanmaktadır. Modern veri analiz dünyasında istatistiğin, bu disiplinlerle entegrasyonu, verinin doğru yorumlanması ve anlamlı sonuçlar çıkarılması açısından vazgeçilmezdir.

İstatistiksel Yazılımlar

İstatistiksel yazılımlar günümüzde istatistik biliminin vazgeçilmez bir parçası haline gelmiştir. Geniş veri setleri üzerinde analiz yapmak, karmaşık modeller oluşturmak ve sonuçları görselleştirmek için bilgisayar yazılımları kullanmak, manuel yöntemlerle mümkün olmayacak kadar kapsamlı analizlerin hızlı ve etkili bir şekilde yapılmasını sağlar. Bu yazılımlar, akademik araştırmalardan endüstriyel projelere kadar her alanda kullanılan çok çeşitli araçlar sunar. İstatistiksel yazılımlar, kullanıcıya modelleme, veri yönetimi ve grafik oluşturma gibi fonksiyonlar sağlar. Bazı yazılımlar ticari olarak sunulurken, bazıları açık kaynaklı ve ücretsizdir. Her yazılımın sunduğu özellikler ve kullanım kolaylığı farklılık gösterebilir; bu da kullanıcıların ihtiyaçlarına göre seçim yapmasına olanak tanır.

Yaygın olarak kullanılan istatistiksel yazılımlar şunlardır:

R: Bu kitapta kullanılan R, açık kaynaklı ve güçlü bir istatistiksel programlama dilidir. R, geniş bir paket yelpazesi ve kullanıcı topluluğu sayesinde istatistiksel analizlerde esneklik sağlar. Veri analizinden makine öğrenimi algoritmalarına kadar birçok alanda kullanılan R, kullanıcıya istatistiksel modelleme, veri görselleştirme ve simülasyon olanakları sunar. R'in popülerliği,

esnekliđi ve topluluk desteđi, onu hem akademik d nyada hem de iř d nyasında  nemli bir ara haline getirmiřtir. Ayrıca, RStudio gibi entegre geliřtirme ortamları, bu dili daha eriřilebilir ve kullanıřlı hale getirir.

SPSS: Sosyal bilimlerde ve pazar arařtırmalarında yaygın olarak kullanılan bu ticari yazılım, kullanıcı dostu bir aray ze sahiptir. SPSS, anket analizleri, regresyon modelleri ve temel istatistiksel testler gibi iřlemler iin idealdir. Kullanıcının programlama bilgisi gerektirmeden analiz yapabilmesi, SPSS'in yaygın olarak tercih edilmesinin bařlıca sebeplerinden biridir.  zellikle anket verilerinin analizi ve sosyal bilimler arařtırmalarında tercih edilen SPSS, sunduđu geniř kapsamlı raporlama ve tablo oluřturma aralarıyla dikkat eker.

Excel: Microsoft Excel, istatistiksel analizler iin kullanılabilecek temel iřlevlere sahip, yaygın bir elektronik tablo programıdır. Excel, b y k veri k melerinin hızlı bir řekilde iřlenmesi ve g rselleřtirilmesi iin ideal bir aratır.  zellikle k  k  lekli analizler ve veri d zenleme iřlemleri iin pratik bir  z m sunar. Kullanıcılar, Excel'de yerleřik fonksiyonlar sayesinde ortalama, standart sapma ve regresyon analizi gibi temel istatistiksel hesaplamaları kolayca yapabilirler. Bununla birlikte, daha karmařık istatistiksel analizler iin sınırlı kalabilir, bu durumda ek eklentiler veya daha geliřmiř yazılımlar tercih edilebilir.

SAS: B y k veri setleri  zerinde hızlı analiz yapabilme kabiliyeti ile bilinen SAS, ticari bir yazılım olarak birok sekt rde yaygın řekilde kullanılmaktadır.  zellikle bankacılık, sađlık ve sigorta gibi b y k veri gerektiren sekt rlerde tercih edilen SAS, g  l  istatistiksel aralar ve veri y netimi  zellikleri sunar. SAS'ın g  l  yanlarından biri de geliřmiř raporlama ve veri madenciliđi yetenekleridir.

Stata: Ekonomi, ekonometri ve sosyal bilimler arařtırmalarında yaygın olarak kullanılan ticari bir yazılımdır. Stata,  zellikle panel veri analizleri, nedensel ıkarım ve zaman serisi analizleri gibi alanlarda g  l  aralar sunar. Kullanıcı dostu komut sistemi sayesinde, Stata, hem basit hem de karmařık istatistiksel analizler iin esnek bir platform sunar. Geliřmiř veri y netimi, modelleme ve grafik oluřturma yetenekleri, onu arařtırma projelerinde yaygın bir seim haline getirmiřtir.

Python: Veri bilimi ve makine  đrenimi alanında yaygın olarak kullanılan Python, istatistiksel analizlerde de g  l  bir aratır. Pandas, NumPy, SciPy ve StatsModels gibi k t phaneler sayesinde Python, hem veri iřleme hem de istatistiksel modelleme aısından b y k esneklik sađlar. Python'un aık kaynak olması ve geniř bir topluluđa sahip olması, onu veri bilimi projelerinde pop ler bir dil haline getirmiřtir.

Bunların dıřında Minitab, JMP, Julia gibi bařka yazılımlar da bulunmaktadır. Her bir yazılım, farklı alanlara ve ihtiyalara y nelik  zel aralar ve  zellikler sunarak kullanıcıların istatistiksel analizleri daha verimli yapmalarını sađlar. Bu nedenle, hangi yazılımın kullanılacađı, analiz edilen verinin t r ne, analiz y ntemlerine ve kullanıcının uzmanlık d zeyine g re deđiřiklik g sterebilir.