

Bilişimde İstatistiksel Analiz

Ders 1

Ali Mertcan KOSE Msc.

`amertcankose@ticaret.edu.tr`

İstanbul Ticaret Üniversitesi



İSTANBUL TİCARET
ÜNİVERSİTESİ

Neden İstatistik?

İstatistik bilimi, büyük miktardaki sayısal verilerin toplanması ve analizinin yapıldığı bir bilimdir.

Neden İstatistik?

1903 yılında, H.G. Wells okuma ve yazma yeteneğinin iyi bir vatandaş olmak için bir gün gerekli olabileceğini istatistiksel olarak düşünerek hipotez olarak kurmuştur. İstatistik çok kısıtlı karar verme süreçlerinde önemli bir rol oynamaktadır. Örnek olarak, yeni bir ilacın pazarlanabilir olmasından önce, Birleşmiş Milletler Gıda ve İlaç yönetimi insan deneklerinin içinde yer aldığı deneysel çalışmalara ve klinik denemelere tabii olmasını gerektirir. ilacın yalnızca etkili olmasını değil aynı zamanda güvenli olup olmamasını belirlemek için veri bu çalışmalardan toplanmak ve analiz edilmek zorundadır.

Neden İstatistik?

Ek olarak, Birleşmiş milletler hükümetinin kararları hakkında sosyal güvenlik ve kamu sağlığı programları ulusal nüfusun uzun ömürlülüğü hakkındaki tahminleri üzerinden meydana gelmektedir; sonuç olarak her bir bireyin yaşayacağı yılların sayısı tahmin edilebilir olmak zorundadır. Bunun gibi, diğer birçok sorunun ele alınması gerekir.

Hükümetlerin bebek ölümlerini azaltmak için nerelere kaynak yatırımı yapması gereklidir? Motorlu araçlardaki kazaların ölüm şansını hava yastığı ve emniyet kemeri kullanımı azaltmakta mıdır? Riski arttıran hangi faktörler bireylerin koroner kalp hastalığını geliştirecektir? Buradaki ve diğer soruların cevabı biyoistatistiksel yöntemlere dayalıdır.

- Biyoistatistik
- Biyometri
- Sosyometri
- Psikometri
- Ekonometri

İstatistiksel yöntemler, işletme, psikoloji ve tarım içeren sayılı alanlarda kullanılabilir. Bu yöntemler biyolojik ve sağlık hizmetleri üzerinde odak noktası gösterdiğinde, Bioistatistik kavramını kullanırız.

İstatistik Verinin toplanması, düzenlenmesi, özetlenmesi ve analiz edilmesi ile ilgili çalışmaların yapıldığı bir alandır.

İstatistik bir veri bilimidir.

- Çok sayıda bilginin özetlenmesini sağlar
- Eldeki az sayıda veriden yola çıkarak öngörü yapar(Bütüne ilişkin çıkarımda bulunur).

İstatistik Nedir?

İstatistiki çalışmalar iki alt düzeyde incelenebilir.

- Betimsel İstatistik
 - Tablolar
 - Grafikler
 - Merkezi Eğilim Ölçüleri
 - Dağılım Ölçüleri
- Tümevarımsal İstatistik
 - Tahmin
 - Kestirim (Bir zaman boyutu vardır)
 - Hipotez testleri (Parametre hakkında iddaa)

Birim (İstatistik Birimi): Üzerinden bilgi derlenen canlı ya da cansız olgulardır.

Örnekleme Birimi: Örnekleme seçilen birime “*Örnekleme Birimi*” denir.

Not

Örnekleme birimi bir ya da birden fazla gözlem birimi içerir.

Kitle: Üzerinden araştırma yaptığımız birimler topluluğuna “*Kitle*” denir.

- **Somut Kitle:** Kitledeki birimlerin bir listesi yapılabiliyorsa “*Somut Kitle*” denir. (Bu listeye çerçeve denir)
- **Soyut Kitle:** Kitledeki birimlerin bir listesi yapılamıyorsa “*Soyut Kitle*” denir.

Not

Eğer sizin kitleniz somutsa, bu kitlenin tamamını ya da alt kümesini seçebilirsiniz. *Kitlenin tamamına ulaşırsa “Tamsayım” alt kümesine ise “Örneklem” dir.*

Somut => Tamsayım, Örneklem

Soyut => Örneklem

Tamsayım => Maliyet, İnsangücü, zaman (Birim zarar görürse olmaz)

Örnekleme: Kitlenin içerisinde örneklem seçme olayına *Örnekleme* denir

Örnekleme: Çekildiği Kitleyi temsil ettiği düşünülen ve evrene göre daha az sayıda birey ya da gözlemden oluşan alt kümedir. Basit bir ifade ile kitlenin bir bölümüne *örneklem* denir. (Kitleyi iyi temsil etmesi gerekir!)

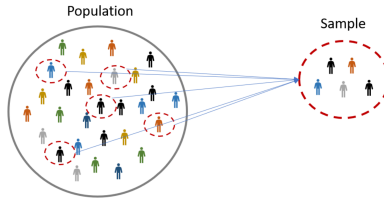


Figure 1: Kitle-Örnekleme.

Parametre: Kitleden hesap edilen bütün değerlere *Parametre* denir.

İstatistik: Örneklemden hesap edilen bütün değerlere *İstatistik* denir.

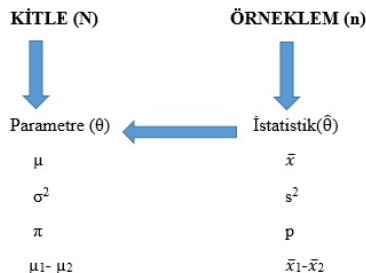


Figure 2: Parametre-İstatistik.

Ölçme: İstatistik biriminin özelliklerine değer verme işlemine *Ölçme* denir.

Değişken: Birimden birime farklılık gösteren özelliklere *Değişken* denir.

Nitel Değişken: Sözcüklerle ifade edilen değişkenlerdir.

Nicel Değişken: Sayılarla ifade edilen değişkenlerdir.

Not

Örneklem uzayından reel sayılara olan bir fonksiyona **rassal değişken** denir.

Kesikli Değişken: Sadece belirli değer alabilen (tamsayı) değişkenlerdir (Günlük tüketilen sigara sayısı, Çocuk sayısı, nabız sayısı gibi).

Sürekli Değişken: İki değişken arasında sonsuz sayı yerleştirdiğimiz değişkenlerdir. Tartarak, ölçerek veya herhangi bir analiz metodu ile elde edilmiştir.(Yaş,Boy, kolesterol seviyesi gibi)

- ① **Sınıflama(Nominal):** Nitel, Sıralama yok; “Cinsiyet”, “Medeni Durumu”
- ② **Sıralayıcı(Ordinal):** Nitel, Sıralama var; “Eğitim Düzeyi”, “Gelir Düzeyi”
- ③ **Eşit Aralıklı (Interval):** Nicel, mutlak sıfır noktası yok, “Sıcaklık”
- ④ **Oran Ölçme Düzeyi:** Nicel, mutlak sıfır noktası var(0 yokluğu gösterir), “Yaş”

Değişkenlerin Ölçme Düzeyi

Veri: Değişkenler bir araya gelince *veri* oluşur.

Veriler iki şekilde elde edilir.

- 1 Araştırmacı tarafından toplanarak (birincil kaynak)
- 2 Biri tarafından kullanılan verilere ulaşarak (ikincil kaynak)

Not

Kesit Verisi: Zamanın belli bir noktasında birden fazla gözlem birimlerinin toplandığı veridir.

Zaman Serisi: Tek bir gözlemden farklı zaman noktalarından toplanan veriye *zaman serisi* denir.

Panel veri: Birden fazla gözlem biriminden farklı zaman noktalarında toplanan veriye *panel veri* denir.