1. HAFTA

Bize İstatistik ne için lazım?

Veriyi ham olarak elde ettikten sonra ham haliyle kullanmamız mümkün olmuyor. Bu verilerin analiz edilmesi gerekiyor. Buna kişisel veri analizi yapmamız lazım. İstatistik bilmek kişisel veri analizini pratikleştirir.

Hangi tip veriyle nasıl çalışılır bilmemiz lazım.

- Tanımlayıcı İstatistik: Verileri daha kolay almamız için, düzenlememiz için, özetlememiz için, analiz edebilmemiz için gerekli çıkarımları yapmamız için argümanlardır.

Medyan: Ortanca değerdir.

1 2 3 4 5 -> 3

12 24 41 51 67 67 85 99 -> (51+67=118:2)59

Mod: En çok tekrar eden sayı. Mesela yukarıdaki sayıda 67'dir. Mod olmak zorunda değil, olmayabilir. Ve mod birden fazla olabilir.(tüm sayılar tekrar ediyorsa hepsi mod olabilir)

Standart Sapma:

Varyans: (yeni piyasaya çıkacak ürünlere varyans kullanmak daha mantıklı) Çeşitlilik. Bir örneğin veya veri kümesindeki veri noktalarının ne kadar yayılmış olduğunun ölçüsü. Formülü: standart sapmanın karesidir.

Range(aralık): En büyük değerle en küçük değerin farkıdır.

Çeyrekler: Yüzdeliklern çeyrekler şeklinde %25'lik dilimler olarak ifade edilmesi. 3 çeyrek değer vardır.(Q1-Q2-Q3) Q2= medyan, ortanca değer. Aynı zamanda Q1 VE Q3 de ortanca değer olabilir.

Q1 Q2 Q3

%25 %25 %25 %25

Çeyrekler arası aralık: Q3-Q1=

Simetrik Dağılım: Veriler simetrik dağılıyorsa mod, medyan ve ortalama değerleri birbirine eşittir. Standart sapmasına göre ortalama değere olan uzaklığı simetrik olarak dağılacaktır. Tahminlerimiz %100 yada çok yakın olacaktır.

Çarpıklık: Pozitif çarpıklık(sağa çarpıklık-medyan değerine göre ortalama değer sağ tarafta kalıyorsa/ortalama moddan büyükse) Negatif çarpıklık(sola çarpıklık-ortalama moddan küçükse)

1. HAFTA

Basıklık: Aykırı değerlerin çokluğuyla alakalı. Aykırı değerler çok olduğu için üstten basılıp aşağıya inmiş gibi gözükür. Ağır kuyruk görüntüsü varsa biz buna ağır kuyruklu, ağır basık dağılım deriz.

Yanlardan sıkılmış gibi yukarıya doğru daha uzun görünüyorsa hafif kuyruklu deriz.

Korelasyon: İki değişkenin birbiriyle ilişkisi varsa biri artarken diğeri de ona doğru yada ters orantılı artıyorsa korelasyon var diyoruz.

Biri artarken diğeri de artıyorsa -> Pozitif Korelasyon

Tam tersi -> Negatif Korelasyon

Bir değişkenin artması diğer değişkenin artmasına herhangi bir şekilde etki etmiyorsa, azalması azalmasına etki etmiyorsa -> Korelasyonsuz Veri

Normal dağılım: Ortalamaya kıyasla baktığımız zaman simetrik bir yapıda sağ ve sol veriler dağılıyorsa, düzgün görünüyorsa.

Aritmetik Ortalamayı Bulmak İçin -> Veriler simetrik dağılıyorsa 2 veri toplanıp 2'ye bölünür. Ortalama, mod yada medyan değeri bulunur.

Her iki tarafında 2 standart sapması varsa toplamda 4 standart sapma yapar. O zaman -2 ile +2 arasında 4 standart sapma vardır. Mesele 1.7'den 1.1 çıkarılırsa 6 kalır yani 60 cm. 4 aralık olduğu için 4'e böldüğümüzde 15 cm çıkacaktır.

Normal dağılım değerleri bize verildiği zaman biz bunu standartlaştırıyoruz.

Parametre dediğimiz şey popülasyonun sayısal özetidir.

İstatistik dediğimiz şey bir örneğin sayısal özeti olmuş oluyor.

Hipotez testi: Ortaya atılan iddiayı istatistiksel yöntemlerle test etme süreci.

Test etmek istediğim hipotez H0 tam tersi iddia ise H1