İSTATİSTİKTE BİLMEMİZ GEREKEN BAŞLICA TERİMLER

1. Veri:

Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır.

2. Nümerik Veriler:

Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır. Örnek olarak, aşağıdaki veri setinde; ofislere göre gelirler gösterilmiştir. Bu kısımda gelir nümerik bir veridir.

Office	Revenue	Revenue
Whalen	$4,400.00	$4,400.00
Hartstein	$13,000.00	$13,000.00
Fay	$6,000.00	$6,000.00
Raphaely	$11,000.00	$11,000.00
Khoo	$3,100.00	$3,100.00
Baida	$2,900.00	$2,900.00
Tobias	$2,800.00	$2,800.00
Himuro	$2,600.00	$2,600.00
Colmenares	$2,500.00	$2,500.00
Mavris	$6,500.00	$6,500.00

3. Nominal (Kategorik) Veriler:

Kategorik bir veri çeşidir. "Daha fazla" ifadesi ile kullanılmazlar.

İkiye ayrılır:

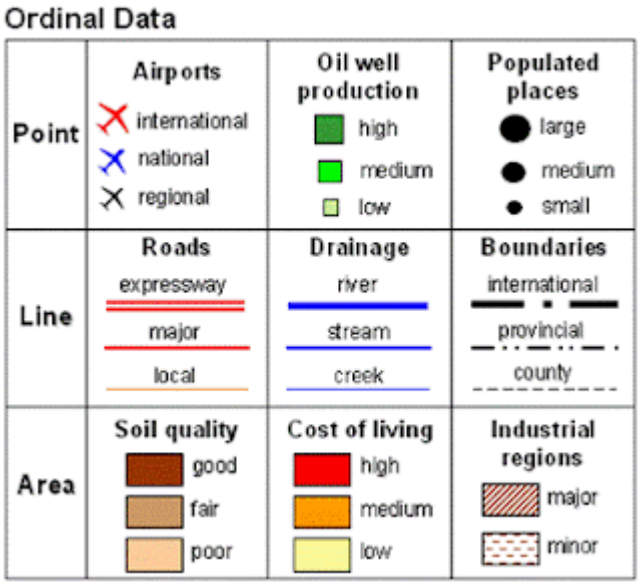
a) Dikotom Veriler : Var-Yok, Kadın-Erkek, Hasta-Sağlıklı
b) İkiden Çok Kategorili : Medeni Durum-Renk-Irk-Şehir, İsim, Forma Numarası

Örneğin forma numarası oyuncunun seviyesi ile ilgili bir bilgi içermez. Yine aynı veri setimizdeki, ofis kısmı nominal verilerden oluşmaktadır.

Office	Revenue	Revenue
Whalen	$4,400.00	$4,400.00
Hartstein	$13,000.00	$13,000.00
Fay	$6,000.00	$6,000.00
Raphaely	$11,000.00	$11,000.00
Khoo	$3,100.00	$3,100.00
Baida	$2,900.00	$2,900.00
Tobias	$2,800.00	$2,800.00
Himuro	$2,600.00	$2,600.00
Colmenares	$2,500.00	$2,500.00
Mavris	$6,500.00	$6,500.00

4. Ordinal Veriler:

Ordinal veriler de yine kategorik veri türündendir. Fakat değerleri arasında sıralı bir ilişki bulunmaktadır. "Daha fazla" ifadesi ile kullanılabilirler ancak ne kadar daha fazla olduğunun ölçüsünü veremezler. Örneğin: Eğitim Düzeyi, Sosyo ekonomik ölçek skorları gibi. Nominal veriler, ordinal verilere daha az bilgi taşırlar.



5. Interval (aralıklı) veriler:

- Nesnelerin sıralanmasında kullanılır.
- Eşit aralıkların eşit mesafelerini temsil ettiği bir ölçek türüdür.
- Ordinal ölçümün bütün özelliklerini taşır.
- Nesneler arasındaki farkın mukayesesine imkan tanır.
- Başlangıç noktası sorunu vardır. Araştırmacı kendi yaptığı çalışmaya göre bir başlangıç noktası belirler.
- Sıfır noktası sabit olmadığı için başlangıç noktası görecelidir.

(*Örnek: Küçük ve büyük tansiyonların değer aralıkları, Kan pH'ı 7.1 ile 7.5 arasında olması, vücut sıcaklığını ölçmek için kullanılan termometreler, likert ölçeği (tutum ve davranışları ölçmek için 1 ile 5 arasında değer vermek gibi, ya da kötü—harika tarzı anlamsal dizilimler*..)

6. Ratio (Oransal) Veriler:

- Interval, ordinal ve nominal ölçüm türünün özelliklerini taşır.
- Aynı zamanda mutlak sıfır noktasına da sahiptir.
- En üst ölçüm tekniğidir, her türlü istatiksel ve matematik işleme imkan tanımaktadır.
- Cevaplayıcı sıfırdan herhangi bir sayıya kadar cevap verebilir.
- Oransallık söz konusudur.

(*Örnek: Yıllık kazancınız ne kadar, kaç çocuğunuz var, sağlık alanında ağırlık ölçümü, boy ölçümü, fizik bilimindeki ölçümler ağırlık, alan, hacim gibi, ölen hastalanan yaralanan hakkındaki veriler vb.*)

7. Değişken:

- Bir değişken karakteristik veya nümerik değer olabilir bu deneyden deneye değişir.
- Bir değişken size bir sayımı (Örneğin; sahip oludğun evcil hayvanların sayısı) veya bir ölçümü yansıtabilir (Örneğin; sabahları uyanma saatlerinizin ölçümü)
- Veya değişkenler her bireyin belirli ölçütlere göre gruplandığı katagorik verilerden oluşablirler (Örneğin; politik görüş, ırk ya da medeni durum)
- Gözlem bilgilerinin, üzerine kaydedilen gerçek veri parçalarıdır.

8. Örneklem:

En basit tanımı ile örneklem; bir evrenin tamamının ölçülemediği durumlarda, evreni en iyi temsil edebileceğine inanılan rassal seçilmiş, yeterli büyüklükteki kümedir.

İstatistikçiler örneklemi anlatırken şu örneğe sıklıkla başvururlar:

> *"Bir kazan çorba düşünün. Denir ki, çorbanın tuzlu olup olmadığını anlamak için kazandaki çorbanın tamamını içmemiz gerekmez. bir kaşık çorba yeter. yeter ki, çorba iyi karıştırılmış ve sizin kaşık çorbayı o bölgeden almış olsun!*
> *eğer çorba iyi karıştırılmamışsa, sizin aldığınız kaşık çorba zehir gibi tuzlu ya da iç bayacak kadar yavan olabilir."*
> *İşte buradaki kazan çorba evrendir, bir kaşık çorba ise örneklemdir.*

Deneysel araştırmaların sağlıklı ve güvenilir gerçekleşmesi için uyun örneklem yöntemi, planlı ve dikkatli hazırlanması gerekmektedir.

Örneklemin evreni temsiliyetini sağlaması için uygun yöntem seçilmesi gerekir. Örneklem evrenin içinden gelen, nitelik ve nicelik açısından onu temsil eden, küçük bir modelidir.

9. Örneklem yöntemlerinden Rastlantısal Seçim:

Araştırmacı örnekleme dahil olacak örnekleri kendi karar vermez. Sadece evrene göre sınırlılıklarını belirtir ve ona göre bu sınırlar içerisinde kalan örnekler -rastgele- olarak seçilir. **Olasılıklı örneklemdir.** Örnekleme seçilen birim-bireylerin örneğe seçilme olasılıkları başlangıçta bellidir. Bu şekilde seçim ile seçime bağlı yanlılık (selection bias) olasılığını azaltmak amaçlanır.

Alt tipleri şunlardır:

1. **Basit rastgele örnekleme**
2. **Sistematik rastgele örnekleme**
3. **Tabakalı rastgele örnekleme**
4. **Küme tipi rastgele örnekleme**

10. Parametre (Parameter):

İstatistik örnek verilere dayanır, popülasyon verilerine değil. Eğer verimizi tüm populasyondan toplamışsak, bu toplama işlemine sayım denir. Eğer biz bu sayımı sayıyla bir değişken üzerinden özetliyorsak , bu sayı istatistik değil parametre olur.

11. Ortalama (Mean or Average):

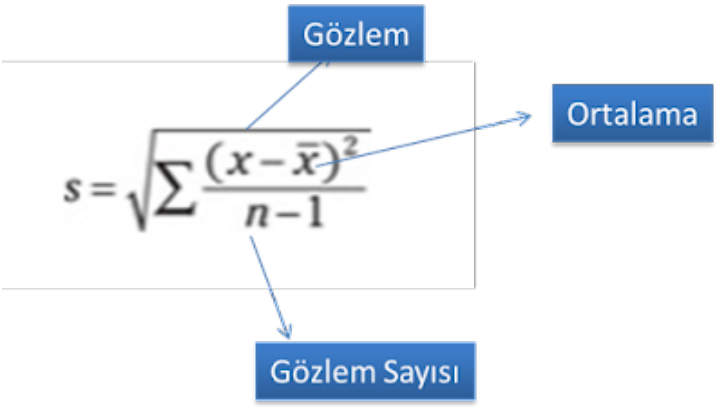
Ortalama istatistikçilerce nümerik verilerdeki, bir ölçümün merkezi veya ortasını bulmak için kullanılmaktadır. Aritmetik ortalama bütün değerlerin toplamının, değer sayısına bölünmesiyle bulunmaktadır. Ortalama her zaman gerçeği yansıtmayabilir, çünkü ortalama aykırı değerlerden kolayca etkilenebilmektedir. (Aykırı değer: Çok küçük veya çok büyük değerler)

12. Medyan (Median):

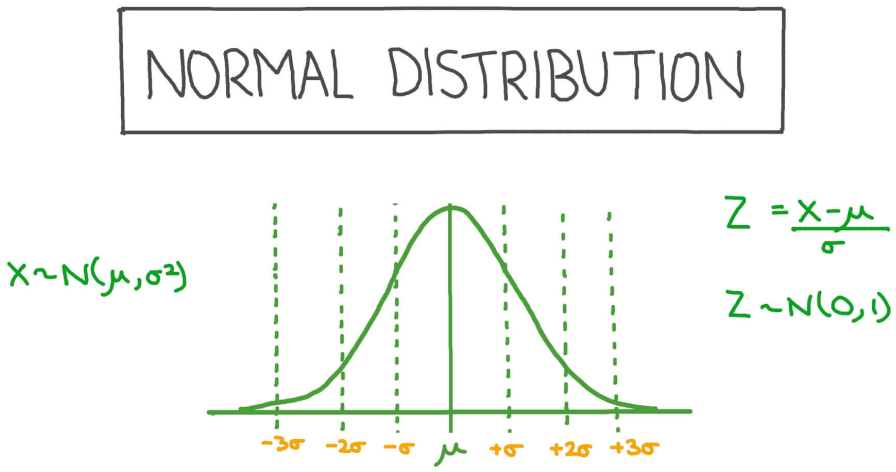
Medyan veri setinin merkezini ölçmek için kullanılan diğer bir yoldur. Basitçe bir veri grubu küçükten büyüğe doğru sıralandığında, ortada bulunan veri ortanca değer (medyan) olarak adlandırılır.

13. Standart Sapma (Standard deviation):

Bir istatistiksel çalışma yaparken beklediğimiz kesin sonuç normalin üstünde veya altında bulunabilir. Zamanla insanlar bu durumu standartlaştırmak istemişler ve bunun için standart sapmayı bulmuşlar. Bu sayede Standart Sapma; yaptığımız çalışmanın sonuçlarının ortalamanın ne kadar üstüne ve altına sapabileceğini hesaplamamıza olanak sağlanmıştır. Standart sapma istatistikçilelerin değişim tutarını ölçmek için kullanıldığı bir ölçüdür. Formülize edersek;



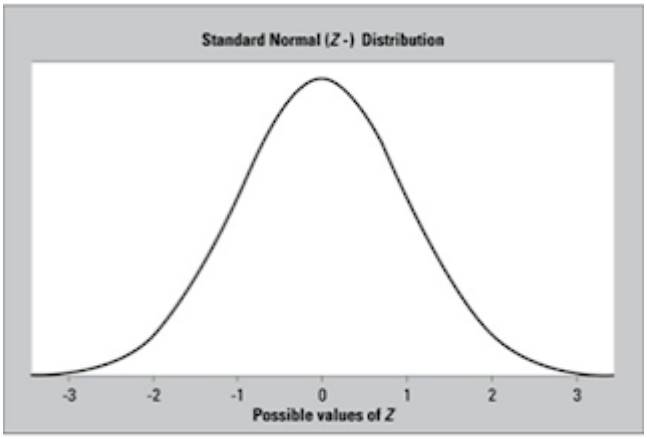
14. Dağılım ve Normal Dağılım (Distribution and normal distribution)



Dağılım bir veri setinin olası değerler listesidir veya tüm olası değerleri veya ne sıklıkla meydana geleceğini gösteren fonksiyondur.

En çok bilinen dağılımlardan biride normal dağılımdır. Normal dağılım olası değerlerin reel sayılar hattına uzanması halinde oluşur ve verinin çoğu (%68 civarı) merkezde ortalama civarindadır. Daha ileri hareket ettiğimiz zaman küçük değerler karşımıza çıkar.

Ortalama her zaman normal dağılımda orta kısımdadır ve standart sapma ortalamadan büküm noktalarına olan uzaklık ölçülerek bulunur (Büküm noktaları ; iç bükeylikten dış bükeyliğe geçişinolduğu noktalar) Bu grafikte ortalama 0 ve standart sapma 1 dir.



15. Merkezi Limit Teoremi (Central Limit Theorem)

N tane birbirinden bağımsız rassal değişkenin (random variable) toplamı olan bir rassal değişkenin n sonsuza giderken ortalama değeri etrafında normal dağılıma sahip olacağını söyler. Daha basit bir ifadeyle birbirinden bağımsız pek çok faktörün etki ettiği bir olayın normal dağılıma yakın davranacağını söyleyebiliriz. Bu sayede ana kütle dağılımı bilinmeyen değişkenler için, örneklemler oluşturarak normal dağılımı kullanabiliriz.

16. Z- Değerleri (Z-values)

Eğer veri setimiz normal dağılıma sahip ise ve bu tüm veri setini bir standart skora standardize ettiğimizde elde edilen skor veri setimiz için bulduğumuz Z-değeri olur. Başka bir tabirle Z skoru ortalamadan kaç standart sapma uzakta olduğumuzu belirtir. Tüm Z değerleri Standart Normal Dağılım olarak bilinen dağılımdan çıkar ve Standard Normal Dağılımda ortalama 0'a Standart Sapma ise 1'e eşittir.

17. Yanılma Payı (Margin of Error)

Muhtemelen şu sözü duymuşsunuzdur: ''Bu analizde yüzde %05 yanılma payı vardır'' Peki nedir bu yanılma payı? Hatasız bir test yapamayacağımız için her testte bir miktar yanılma riskimiz vardır. Bunu 0,05 ; 0,01 ; 0,005 ; 0,0001;... gibi bir düzey olarak benimseyebiliriz. Yanılma payımız küçüldükçe, teste olan güven düzeyimiz yükselir. O nedenle istatistikçiler olabildiğince az yanılma ile test yapmak isterler. Yine de α =0,05 ve α=0,01 düzeyleri en çok kullanılanlardır.

18. Güven Aralıkları (Confidence interval)

Şimdi Türkiye geneli hane geliri araştırması yaptığımızı düşünelim. Bunun için popülasyonumuz ''Tüm Türkiye'nin hane geliri '' ve parametremizi '' Ortalama Hane Gelir ''olarak varsayalım. Başlarken populasyondan bir örneklem aldık (Örneğin Türkiye hanelerinden 1000 tanesi) Ve sonra bu örnekleme ilişkin istatistiklerimizi bulduk. Eğer tüm bir popülasyon hakkında bir yorum yapmak istiyorsak bunun sonrasında bulduğumuz değerlere basit bir tabirle ekleme veya çıkartma yapmamız gerekir çünkü örnekten örneğe değişim görülebilir. Bu ekleme ve çıkama, örneklem istatisğinin tahmini amacıyla eklediğimiz parametrenin yanılma payıdır. Ve bunun sonucunda güven aralığımızı elde ederiz. Basit bir Örnekle evden işe 30 dkda gidiyorsunuz ama bu bazen 5 dk erken bazen 5 dk geç oluyor burdaki 5 dk sizin yanılma payınız ve güven aralığınız ise 25 ve 35 arasındadır.

19. Hipotez Testi (Hypothesis testing)

Hipotez kısaca doğruluğu bir araştırma ya da deney ile test edilmeye çalışılan öngörülere, denencelere denir.

Hipotez testleri bir örneklem ortalaması ile bu örneklemin çekilmiş olduğunu düşündüğümüz ortalaması etrafındaki farkın anlamlı olup olmadığını (yani önemli bir fark olup olmadığını) araştırmamızı sağlayan testlerdir.

Örneğin; Bir pizza şirketi ortalama siparişlerini adreslere 30 dakikada teslim ediyor ve bize bunu test etmemiz isteniyor. Bunun için tüm sipariş sürelerini içeren ana kütleden rastgele örneklem seçimi yaprak ortalama sürenin ortalama periyodun dışına çıkıp çıkmadığını kontrol edip bu iddianın anlamlı olup olmadığını test ederiz. Ayrıca bulduğumuz sonuçlar örneklemden örnekleme farklılık gösterebilceğinden bu hesap değerlerinide dikkate alınmalıdır.

Örnek:

H_0 : Ortalamalar arasında anlamlı bir farklılık yoktur.

H_1 : Ortalamalar arasında anlamlı bir farklılık vardır.

20. p – Değeri ve Anlamlılık (p-values)

Hipotez testi popülasyon hakkında yapılmış olan iddianın gerçekliğini test etmek için kullanılır. Bu denemede olan iddiaya sıfır hipotezi dersek ve bu hipotezin asılsız sonuca varılması durumunda inanacağımız diğer hipotez ise alternatif hipotez olur. Hipotez testlerinin doğruluğunun gücünü ölçmek için p – değerini kullanırız. P değeri 0 ile 1 arasında olan bir numaradır.

Küçük "p" değeri (p <= 0.05) sıfır hipotezini reddederiz.

Büyük "p" değeri (p>0.05) sıfır hipotezini reddedemeyiz.

Örneğin biraz önceki örnekten gidelim; bir Pizza şirketi siparişlerini evlere ortalama en fazla 30 dakikada getirdiğini söylüyor, bizde buna inanmıyoruz ve 30 dakikadan fazla olduğunu iddia ediyoruz.Buna göre hipotezimiz;

H_0 : Eve servis süresi max. ortalama 30 dk.

H_1 : Eve servi süresi 30 dk.dan daha fazladır.

Rastgele eve servis sürelerinden örneklem seçtik ve hipotez testi kanalıyla verimizde gerekli işlemleri yaptık ve p-değerimiz 0,001 olarak çıktı. Ve bu değer 0.05 den küçük. Bu durumda sıfır hipotezini reddettik ve alternatif hipotezi kabul ettik. Yani iddiamızın doğruluğunu kanıtlamış olduk.

Linkler:

Link 1, Link 2, Link 3, Link 4, Link 5, Link 6, Link 7