İstatistik Nedir?

“İstatistik” rasgelelik içeren olaylar, süreçler, sistemler hakkında modeller kurmada, gözlemlere dayanarak bu modellerin geçerliğini sınamada ve bu modellerden sonuç çıkarmada gerekli bazı bilgi ve yöntemleri sağlayan bir bilim dalıdır.

Sistem-Model Kavramı

Gerçek dünyadaki bir olayın, sürecin veya birimlerden oluşan ve birimleri arasındaki iç ilişkiler yanında çevre ile dış ilişkilere göre işleyen bir **sistemin** belli bir anlatımına **model** denir. Anlatım sözle, çizimle, belli bir ölçekte fiziki benzer oluşturmak veya başka bir şekilde yapılmakla birlikte en geçerli anlatım, bilimin ortak dili olan matematik ile yapılmaktadır.

Model, gerçek dünyadaki bir olgunun belli bir anlatımı olmak üzere, **simülasyon**, model üzerinde "deney yapmaktır". Sistem, belirli girdileri olan ve bunları işleyerek bilgi veren elemanlar topluluğu olarak tanımlandığı gibi, birbirleriyle etkileşimli olan elemanların sıralanmış bir kümesi olarak da tanımlanabilir.

Model

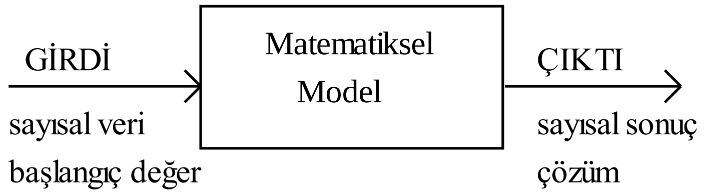
Model gerçek dünyadaki bir olgunun veya sistemin yapı ve işleyişinin, ilgili olduğu bilim sahasının (fizik, kimya, biyoloji, jeoloji, astronomi, ekonomi, sosyoloji, vb.) kavram ve kanunlarına bağlı olarak ifade edilmesidir. Model gerçek dünyadaki bir olgunun bir anlatımıdır, bir temsilidir. Gerçek dünyanın çok karmaşık olması nedeniyle modeller, anlatmak istedikleri olgu ve sistemleri basitleştirerek belli varsayımlar altında ele almaktadır. Modeller gerçeğin kendileri değildirler ve ne kadar karmaşık görünseler de gerçeğin bir eksik anlatımıdırlar. Kısaca model denilen şey; model kurucunun gerçeği "anlayışının" bir ürünüdür.

Modeller değişik biçimlerde sınıflandırılmakla birlikte en geçerli anlatım biçimi matematiksel modellerdir. Matematiksel modeller;

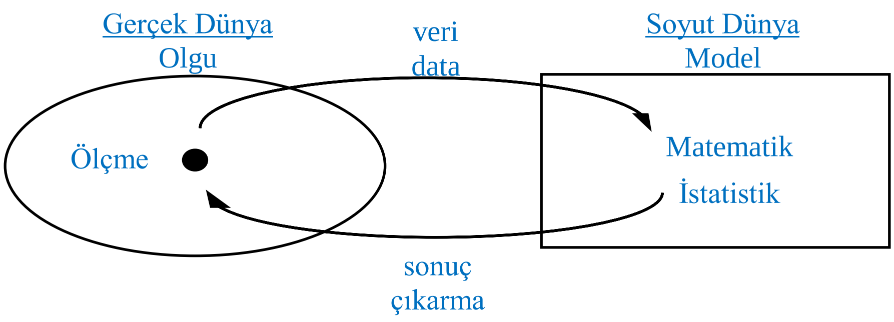
*Stokastik (rasgele değişken içeren) ve deterministik (rasgele değişken içermeyen) matematiksel modeller,

*Doğrusal ve doğrusal olmayan modeller,

*Sürekli (diferensiyel denklem,...) ve kesikli (fark denklemi,...) modeller olmak üzere sınıflandırılabilir. Matematiksel modeller anlatım gücü en fazla ve en geçerli olan modellerdir. Genel olarak bir matematiksel model aşağıdaki gibi ifade edilir.



Gerçek dünyayı anlama ve anlatmada, yani modellemede insan aklının en güçlü iki aracı matematik ve istatistiktir. İstatistik özellikle, rasgelelik içeren olguların modellenmesinde ön plana çıkmaktadır. Herhangi bir deneysel bilimin ilgi sahasına giren olguları modellemede düşünce tarzı aşağıdaki gibidir.



Betimsel İstatistik

Verilerin toplanması ve bu verilerden sonuç çıkarılması bir istatistiksel araştırmanın en önemli kısımlarını oluşturmaktadır. Bu veriler, bir deney yoluyla elde edilebileceği gibi anket ve gözlem yoluyla da elde edilebilir. Günümüz teknolojisi sayesinde veri toplamak ve bu verileri saklamak kolay hale gelmiştir. İstatistiksel çalışmada amaç, toplanan çok büyük miktardaki veriden bilgi çıkarmaktır.

Betimsel istatistikler, bir çalışmadaki verilerin temel özelliklerini anlatmak, tarif etmek, tasvir etmek ve ifade etmek için kullanılan yöntemleri içerir.

Betimsel istatistik yöntemleri toplanan bu verileri, basit-kullanışlı ve anlaşılabilir bir biçimde özetlemek amacıyla, çeşitli grafik ve çizelgeler kullanırlar. Bu istatistiklerin her biri veriyi sadece betimler. Eldeki verilerin altında yatan süreçlerin ortaya çıkarılması amacını içermezler. Tüm bu sınırlılıklarına rağmen konu ile ilgili kimselerin karşılaştırma yapabilmesine olanak tanırlar.

Çıkarımsal İstatistik

Çıkarımsal istatistikle, araştırma için toplanan verilerin içinde gizli kalmış anlamların ortaya çıkarılması amaçlanır. Eldeki verilerden elde edilen bilgiler, geleceğe ilişkin kestirimlerde bulunmak, sonuç çıkarmak, tahmin yapmak ve karar vermek için kullanılır. Burada asıl sorun, bu verilerden elde edilen sonuçlara dayanarak genel geçer genellemenin ne derecede yapılabileceğidir.

VERİLERİN DÜZENLENMESİ VE ÖZETLENMESİ

Rasgelelik içeren olgulardan elde edilen ölçüm (gözlem) değerlerine **istatistiksel veri** veya kısaca **veri** denildiği belirtilmişti. Veriler, deneyler sonucu veya doğal koşullarda olguları gözlemlemekle elde edilir. Rasgelelik söz konusu olduğunda, ölçme sonucunda çıkan sayılar da rasgele olacaktır. Bu rasgele sayılar, bir **rasgele değişkenin** aldığı değer olarak düşünülmektedir. Rasgele değişkenlerin bir **dağılımı** vardır. Bu dağılımların biçimlerinin görülebilmesi için, ölçümlerden elde edilen verilerin anlaşılabilir bir şekilde sunulmaları, betimlenmeleri gerekmektedir. Gözlem sayısı çok olduğunda bu gözlemler çeşitli çizelgeler ve grafiklerle sunulur. Bu şekilde veriler düzenlenir ve özetlenir.

Sıklık Çizelgeleri
Histogram
Birikimli Sıklık Grafiği
Gövde Yaprak Grafiği
Kutu Grafiği
Saçılım (Serpme) Grafiği

Değişim Genişliği (R): Örneklemdeki en büyük değer ile en küçük değer arasındaki farka değişim genişliği denir.

$$R = En\ büyük\ gözlem\ değeri - En\ küçük\ gözlem\ değeri$$

n birimlik bir örneklemin sıklık çizelgesini oluşturabilmek için ilk olarak değişim genişliği R hesaplanmalıdır. Daha sonra, bu R, eşit uzunluktaki aralıklara bölünerek sınıflar elde edilir.

Sınıf Aralığı: n birimlik bir rasgele örneklemin değişim genişliği R'nin bölündüğü aralıklara sınıf aralıkları denir. Sınıf aralığı c ile gösterilmek üzere;

$$c = \frac{Değişim\ genişliği\ (R)}{İstenilen\ sınıf\ sayısı\ (k)}$$

ile hesaplanır.

Alt Sınır (As): Bir sınıf aralığının en küçük değeridir.

Üst Sınır (Üs**):** Bir sınıf aralığının en büyük değeridir.

Sınıf Ortası (S_i): Bir sınıf aralığının merkezine ya da orta noktasına sınıf ortası denir. Bir sıklık çizelgesinde k tane sınıf aralığı varsa sınıf ortaları $S_1, S_2, \ldots, S_k$ ile gösterilir.

Sıklık (Frekans-f_i): Bir sınıf aralığına düşen veri sayısına sıklık (frekans) denir. Bir sıklık çizelgesinde k tane sınıf aralığı varsa sınıf sıklıkları $f_1, f_2, \ldots, f_k$ ile gösterilir. Sıklıkların toplamı, toplam veri sayısını verir, yani, $\sum_{i=1}^{k} f_i = n$ dir.

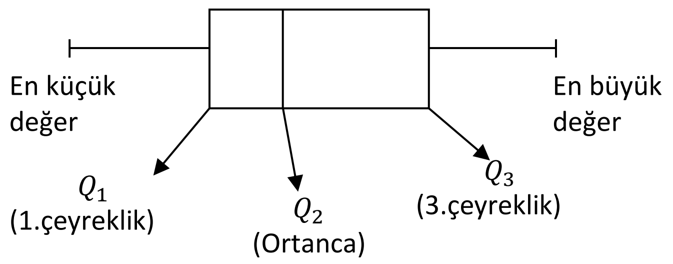
Göreli Sıklık (Frekans Yüzdesi-p_i**):** Her sınıfa düşen sınıf sıklıklarının toplam sıklığa oranıdır. Bir sıklık çizelgesinde k tane sınıf aralığı varsa sınıflar için göreli sıklıklar $p_1, p_2, \ldots, p_k$ ile gösterilir ve $p_i = \frac{f_i}{n}, i = 1,2, \ldots, k\ olmak\ üzere \sum_{i=1}^{k} p_i = 1\,'dir.$

Histogram

Histogram, koordinat sisteminde, tabanları x ekseni üzerinde sıklık çizelgesi tablosundaki her bir sınıfın sınıf aralığı büyüklüğünde, yükseklikleri bulunduğu sınıfın sıklıkları ile orantılı olarak yan yana çizilen dikdörtgenlerden oluşur.

Kutu grafiği

Bir veri grubundan elde edilen en küçük değer, en büyük değer, birinci çeyreklik (Q_1 ile gösterilir, verilerin %25'i bu değerin altında değer alır), üçüncü çeyreklik (Q_3 ile gösterilir, verilerin %25'i bu değerin üstünde değer alır) ve ortanca (Q_2 ile gösterilir, verilerin %50'si bu değerin altında, %50'si bu değerin üstünde değer alır) değerlerini içeren aşağıdaki gibi grafiktir.



KONUM ÖLÇÜLERİ

Aritmetik Ortalama

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Geometrik Ortalama

$$G = \sqrt[n]{x_1 . x_2 \ldots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Harmonik Ortalama

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Tepe Değeri (Mod)

Verilen bir veri grubunda en çok tekrar eden değer tepe değer (mod) olarak adlandırılır.

Ortanca (Medyan)

Gözlem değerleri, küçükten büyüğe göre sıralandığında, tam ortaya düşen değer ortanca değer olarak adlandırılır.

Çeyreklikler

Gözlemler küçükten büyüğe doğru sıralandığında gözlem değerlerini dört eşit parçaya bölen değerlere çeyreklikler denir. Birinci çeyreklik (Q_1), gözlemler küçükten büyüğe sıralandığında gözlemlerin %25'ini solunda, %75'ini sağında bırakan değerdir. İkinci çeyreklik (Q_2), gözlemlerin ortancasına denk gelmektedir. Üçüncü çeyrek değer (Q_3), gözlemler küçükten büyüğe sıralandığında gözlemlerin %75'ini solunda, %25'ini sağında bırakan değerdir. Yani sıralı gözlemlerde, ortancadan küçük olan değerlerin ortancası birinci çeyreklik, ortancadan büyük olan gözlemlerin ortancası üçüncü çeyrekliktir.

Değişim Aralığı

$$R = En\ büyük\ değer - En\ küçük\ değer$$

Varyans

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Standart Sapma

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Standart Hata

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Çeyrekler Arası Sapma

$$Q = \frac{Q_3 - Q_1}{2}$$

Mutlak Sapma

$$ms = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$$

Değişim Katsayısı

$$DK = \frac{s}{\bar{x}} \times 100$$

Eğer bir dağılımda, aritmetik ortalama, tepe değer ve ortanca eşitse dağılımın **simetrik** olduğu söylenir. Aritmetik ortalama ortancadan küçük o da tepe değerden küçükse dağılım **sola çarpıktır** denir. Tepe değer ortancadan küçük o da aritmetik ortalamadan küçükse dağılım sağa çarpıktır denir.

OLASILIK

Rasgele Sonuçlu Deney: Sonuçlarının kümesi belli olan, ancak hangi sonucun ortaya çıkacağı önceden söylenemeyen bir işleme **Rasgele Sonuçlu Deney** veya kısaca **Deney** denir.

Örnek Uzay: Bir deneyin tüm olabilir sonuçlarının kümesine **örnek uzay** denir. Genellikle S harfi ile gösterilir. Örnek uzay, üzerinde çalışılan evrensel kümeye denk gelmektedir.

Olay: Örnek uzayın bir altkümesine **olay** denir. Yalnız bir öğeden meydana gelen alt kümeye basit (ilkel), birden fazla öğeden meydana gelen alt kümeye birleşik olay denir.

Ayrık Olay: İki olay aynı anda meydana gelemiyorsa bu olaylara **ayrık olaylar** denir. Kesişimleri boş küme olan olaylara **ayrık olaylar** denir.

Bir Olayın Olasılığı

Olasılık; tanım kümesi S örnek uzayın tüm alt kümelerinin kümesi ve görüntü kümesi [0,1] arasındaki reel sayılar olan bir fonksiyondur. Olasılık fonksiyonu, örnek uzayı ile olasılık değerleri arasında fonksiyon tipinde bir ilişki kurar ve her olaya, o olasılığı belirten 0 ile 1 arasında bir sayı karşılık getirir. Bir deney yapıldığında, tüm olanaklı sonuçların sonlu S örnek uzayı oluşturulabilir. Örnek uzayı $S = \{D_1, \dots, D_n\}$ ve örnek uzayın tüm alt kümelerinin kümesi K olmak üzere, bir A olayının olasılığı

$$P: K \to R$$
$$A \to P(A)$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{A'nın\ \ eleman\ sayısı}{S'nin\ \ eleman\ sayısı}$$

olarak verilir. Bu tanım **olasılığın klasik tanımıdır**. Bu durumda, örnek uzayın her bir elemanı için $P(D_1) = \cdots = P(D_n) = \frac{1}{n}$ olduğu kabul edilmiş olur. Klasik tanım, örnek uzayının sonlu sayıda birimden oluşması ve her olayın eş olasılıkla ortaya çıkması karşılığında kullanılabilir. Gerçek yaşamda eş olasılıklı olaylarla nadir karşılaşılır.

Olasılık Aksiyomları

Bir deneyin örnek uzayı S olsun. Örnek uzayın tüm alt kümelerinin kümesi K olsun. Her bir A olayı için, $P(A)$ gerçek değerli fonksiyonu "A" olayının ortaya çıkma olasılığı olarak tanımlanır ve aşağıdaki özellikleri sağlar. Yani aşağıdaki özellikleri sağlayan $P(A)$ fonksiyonuna **olasılık fonksiyonu denir.** $P(A)$ değerine de A olayının olasılığı denir.

$$P: K \to R$$

$$A \to P(A)$$

olmak üzere aşağıdaki özellikleri sağlayan fonksiyona **olasılık fonksiyonu** denir. Bu özellikler **olasılık aksiyomları** (Kolmogorov Aksiyomları) olarak bilinir.

A.1) Bütün A olayları için $P(A) \geq 0$'dır. Yani olasılık pozitif veya sıfır olan, negatif değerler almayan bir sayıdır.

A.2) $P(S) = 1$. Kesin olayın olasılığı 1'e eşittir.

A.3) Sayılabilir sayıda aynı anda imkânsız olayın birleşiminin olasılığı, olasılıkları toplamına eşittir. $A_1, A_2, \dots, A_n$ olaylar olsun.

$$A_i \cap A_j = \emptyset \ \ i \neq j \ \ i,j = 1,2,\dots,n \text{ ise}$$

$$\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

$$A_i \cap A_j = \emptyset \ \ i \neq j \ \ i,j = 1,2,\dots \text{ ise}$$

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

olacaktır.

S ve $P(A)$ ya birlikte **olasılık uzayı** adı verilir. S sonlu, sayılabilir sonlu ve sonsuz elemanlı olabilir.

Teorem:

1. $P(\emptyset) = 0$

2. Bir olayın tümleyeninin olasılığı

$$P(A') = 1 - P(A)$$

şeklindedir.

3. Bir olasılık uzayında A ve B kümeleri arasındaki farkın olasılığı, yani $A - B = A \cap B'$ nin olasılığı

$$P(A - B) = P(A \cap B') = P(A) - P(A \cap B)$$

şeklindedir.

4. Bir olasılık uzayında A ve B gibi iki olayın birleşiminin, $A \cup B$ olasılığı

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

şeklindedir.

5. Bir alt olayın olasılığı A ve B aynı olasılık uzayında tanımlanmış olaylar ve $A \subset B$ olsun, bu durumda,

$$P(A) \leq P(B)$$

şeklindedir.

6. Bir olasılık uzayında tanımlanmış tüm A olayları için,

$$0 \leq P(A) \leq 1$$

şeklindedir.

Tanım: S sonlu bir örnek uzay olsun, $S = \{a_1, a_2, \dots, a_n\}$ alalım. Sonlu olasılık uzayı her bir $a_i \in S$ için a_i nin olasılığı denen p_i reel sayısı karşılık getirilerek bulunur. Öyle ki

i. Her $p_i \geq 0,\ i = 1,2, \dots n$

ii. $\sum_{i=1}^{n} p_i = 1$

olacaktır.

Tanım: S sayılabilir sonsuzlukta bir örnek uzay olsun, $S = \{a_1, a_2, \dots, a_n, \dots\}$ alalım. Sonlu uzaydaki gibi, her bir $a_i \in S$ için a_i nin olasılığı denen p_i reel sayısı karşılık getirilerek bulunur. Öyle ki

i. Her $p_i \geq 0,\ i = 1,2, \dots$

ii. $\sum_{i=1}^{\infty} p_i = 1$

olacaktır.

Çarpım Kuralı

Bir deneyde, ilk işlem N_1 yolda yapılabiliyorsa ve işlem bu yolların herhangi birinde yapıldıktan sonra, ikinci işlem N_2 yolda yapılabilirse, bu iki işlem birlikte $N_1.N_2$ yolda yapılabilir (buna **çarpım kuralı** denir).

Nesnelerin kümesinin bir kısmının ya da tümünün belli bir sıralamasına veya düzenlenmesine **permütasyon** denir.

n nesnenin bir grubundan bir defada alınan r nesnenin bir sıralamasına, bu n nesnenin permütasyonu denir. Böyle permütasyonların toplam sayısı $r \leq n$ olmak üzere $P(n,r)$ ile gösterilir ve

$$P(n,r) = \frac{n!}{(n-r)!}$$

ile hesaplanır. Bu, bir defada r tanesi alınarak ve yinelemeden (kullanılan yine kullanılmayacak) n farklı nesnenin **permütasyonlarının** sayısıdır.

Kombinasyon

Tanım: Bir defada r tanesi alınan n farklı nesnenin bir kombinasyonu, düzenleme sırasına bakılmaksızın n nesneden r tanesinin bir seçimidir. Bir defada r tanesi alınan n nesnenin kombinasyonlarının sayısı $C(n,r)$ ile veya $\binom{n}{r}$ gösterilir ve

$$C(n,r) = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

şeklinde hesaplanır. r nesnenin her bir kombinasyonun $r!$ yolda düzenlenebileceğinden, yani $\binom{n}{r}$ kombinasyonlarının her biri için $r!$ permütasyon olduğundan permütasyonların toplam sayısı $\binom{n}{r}.r!$ dir. Bu nedenle, aşağıdaki eşitlik elde edilir.

$$\binom{n}{r}.r! = P(n,r) = \frac{n!}{(n-r)!}$$

Ayrıca, bir defada $n-r$ tanesi alınan n farklı nesnenin kombinasyonlarının sayısı, bir defada r tanesi alınan n farklı nesnenin kombinasyonlarının sayısının aynıdır.

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \binom{n}{n-r}$$

Koşullu olasılık, bir olayın başka bir olayın meydana gelmesi koşulu altında ortaya çıkması olasılığıdır. B olayı bilindiğinde A olayının ortaya çıkması olasılığı,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \;\; P(B) \neq 0$$

ile gösterilir ve B olayı verilmişken A olayının **koşullu olasılığı** olarak adlandırılır. Benzer şekilde, A olayı verilmişken B olayının koşullu olasılığı,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \;\; P(A) \neq 0$$

olarak tanımlanır. Ayrıca,

$$P(A \cap B) = P(A|B).P(B)$$

ve

$$P(A \cap B) = P(B|A).P(A)$$

olarak yazılabilir.

Olasılığın Sıklık Tanımı

Deneme sayısı N ile, A olayının N denemede gerçekleşme sayısı n ile gösterilsin. N sayısı giderek artarken n/N oranı $P(A)$ gibi belirli bir değere yakınsıyorsa, daha matematiksel bir ifadeyle ε istenildiği kadar küçük olabilen pozitif bir sayı olmak üzere, $N > M(\varepsilon)$ iken

$$\left|\frac{n}{N} - P(A)\right| < \varepsilon$$

olmasını sağlayacak bir $M(\epsilon)$ sayısı bulunabiliyorsa

$$\lim_{n \to \infty} \frac{n}{N} = P(A)$$

yazılır ve $P(A)$, A olayının gerçekleşmesinin olasılığı olarak alınır.