

BÖLÜM 1

İSTATİSTİK ve FREKANS DAĞILIMI

1.1. İstatistik Nedir?

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

İstatistik; tüm bilim dallarında, çalışma alanlarında bir amaca yönelik, sayısal değerleri derleme, özetleme, tablolar ya da grafikler biçiminde düzenleme, çözümlenme, sonuçları yorumlama, parametre tahmini yapma, değişkenler arasındaki ilişkinin derecesini belirleme, örneklem arasındaki ayrıcalığı belirleme, deney düzenleme gibi konuları kapsayan bir bilim dalıdır.

1.2. Temel Kavramlar

Kitle: Araştırma kapsamında alınan, aynı özelliği taşıyan birim ya da bireylerin oluşturduğu topluluğa kitle denir.

Örneklem: Örneklem yöntemlerinden yararlanılarak bir kitleden seçilen, aynı özellikleri taşıyan ve kitleyi temsil edebilecek nitelikteki ve nicelikteki bireylerin oluşturduğu topluluğa örneklem denir. Örneklem seçmek için kullanılan yöntemler topluluğuna “örneklem” denir.

Parametre: Kitlenin özelliklerini belirleyen sayısal değerlerdir

Kitle ortalaması: μ , Örneklem ortalaması: \bar{x}

Kitle varyansı: σ^2 , Örneklem varyansı: s^2

Örneklem ortalaması

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad n: \text{örneklem hacmi}$$

Örneklem varyansı

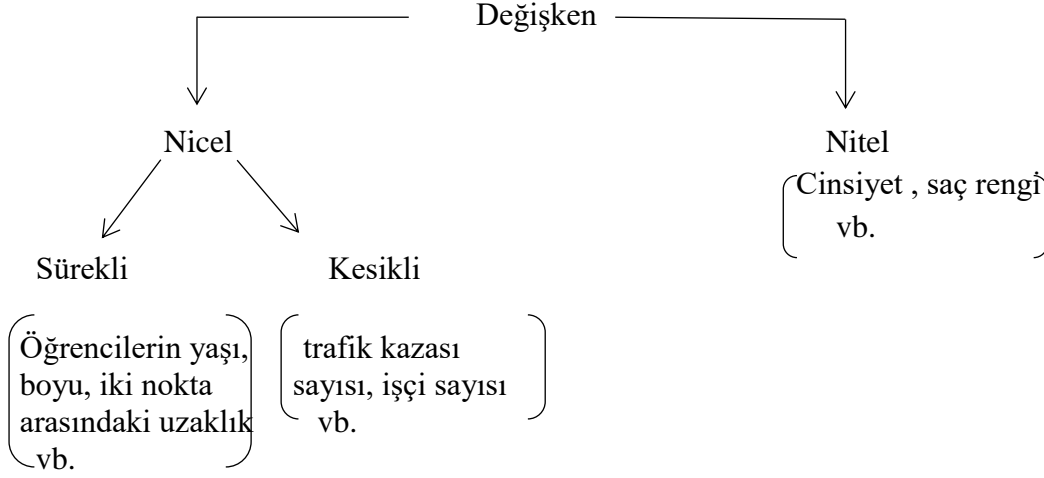
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Değişken: Birim ya da bireylerin her bir özelliğine değişken denir. Değişkenler; $X, Y, Z...$ gibi büyük harflerle, aldıkları değerler ise $x, y, z...$ gibi küçük harflerle gösterilir. Bir değişken sayısal değerlerle ölçülüyorsa “nicel değişken” denir. Nicel değişkenler “kesikli” ve “sürekli” değişken olmak üzere ikiye ayrılırlar.

Kesikli değişkenler genellikle sayılabilir değişkenlerdir. Belli bir aralıkta her değeri almayan değişkenlerdir. Örneğin; öğrenci sayısı, raftaki kitap sayısı...

Sürekli değişkenler ise belli bir aralıkta her değeri alabilen değişkenlerdir. Örneğin; ağırlık, gelir düzeyi, boy uzunluğu...

Nitel deęişkenler ise sayısal deęerlerle ölçülemeyen deęişkenlerdir. Örneęin; göz rengi, cinsiyet vb.



Veri: Üzerinde çalışılan birim ya da bireylerin her birinin aldığı deęerlere veri denir. Veriler iki grupta toplanır; nitel veriler, nicel veriler.

Nicel veriler de kendi içerisinde kesikli ve sürekli olmak üzere ikiye ayrılır. Nitel veriler ise nitelik belirten deęişkenlere ilişkin verilerdir.

Veri

- Gözlem
- Laboratuvar
- Anket
- Kayıt ve arşivler

vb. şekillerde elde edilebilir

İstatistik, teorik istatistik ve uygulamalı istatistik olarak iki başlık altında toplanır.

Uygulamalı istatistik betimsel istatistik ve çıkarımsal istatistik olarak iki alanda ele alınır.

Betimsel İstatistik: Verinin tanımlanması ve özetlenmesi.

Çıkarımsal istatistik: Örneklemeden elde edilen bilgilere dayanarak kitle hakkında tahminlerde bulunmak, sonuçlar çıkarmak ve kararlar vermek.

1.3. Ölçme Düzeyleri

İstatistiksel araştırmalarda, deęişkenlerin ölçülmesi önemli yer tutar. Ölçüm yapılırken kullanılan ölçüm düzeyleri dört başlık altında ele alınabilir.

1. Sınıflama Ölçme Düzeyi: Birimlere özelliklerine göre isimler verilir. Ölçme, isimlendirilerek (sınıflandırılarak) yapılır, herhangi bir sıralama yapılmaz. Örneęin cinsiyet

değişkeni kadın, erkek gibi iki sınıfta ifade edilir. Tedavide kullanılan ilaç türü, medeni durum, meslek, gibi değişkenlere ait ölçümler sınıflama ölçme düzeyine örnek olarak verilebilir.

2. Sıralama Ölçme Düzeyi: Sıralama ölçme düzeyinde değişkenlerin aldığı değerler önem derecesi ya da üstünlüklerine göre sıralanır. Katılım düzeyi (Kesinlikle Katılıyorum, Katılıyorum, Kararsızım, Katılmıyorum, Kesinlikle Katılmıyorum), sıklık düzeyi (Hiç, Nadiren, Genellikle, Her Zaman), öğrenim durumu (İlköğretim, Lise, Lisans, Yüksek Lisans), vb. değişkenler için sıralama ölçme düzeyi kullanılır.

3. Eşit Aralıklı Ölçme Düzeyi: Sıcaklık, başarı, performans gibi nicel değişkenleri ölçmek için kullanılır. Bu ölçekte bir başlangıç noktası bulunmaz. Yani, “0” değeri eşit aralıklı ölçme düzeyinde yokluk ifade etmez. Örneğin termometrede görülen “0 C” belirli bir anlam taşır.

4. Oranlama Ölçme Düzeyi: ekonomik durum , ağırlık, uzunluk, hız, not gibi değişkenleri ölçmek için kullanılır. Bu ölçme düzeyinde başlangıç “0” noktasıdır. Oranlama ölçme düzeyinde yer alan “0 Kg” bir yokluk ifadesidir.

Sınıflama ve sıralama ölçme düzeyi nitel değişkenler için, eşit aralıklı ve oranlama ölçme düzeyi nicel değişkenler için kullanılır.

Aralık veya oranlama düzeyinde yapılan ölçümlerden elde edilen veriler birer niceliktir (sayıdır). Bazen sınıflama düzeyinde yapılan ölçümlerden elde edilen veri de sayılardan oluşabilir. Örneğin para atışında, yazı gelişi 0, tura gelişi 1 olarak kodlanırsa elimizdeki veri sayısal bir veri olacaktır. Tersine de olabilir. Sayısal olarak yapılan ölçümlerden niteliksel veri elde edilebilir. Tavla zarı üzerindeki noktalar sayıldıktan sonra ölçümler tek-çift olarak nitelendirilebilir.

1.5. Frekans Tabloları

1.5.1. Nicel verilerde Frekans tablosu

$k = \text{Sturges kuralı}(\text{Sınıf sayısı}) = 1 + 3.3 \log n$, n :örneklem hacmi

Örnek: Bir endüstriyel bölgede günlük hava kirliliği ölçümleri yapılmış ve sülfüroksit miktarları aşağıdaki gibi bulunmuştur.

Tablo 1.1.

15.8	14.4	22.7	10.5	24.6	14.7	22.3	20.4	24.8	24.3
21.3	8.3	15.2	17.3	19.4	20.5	27.5	16.2	26.1	11.8
27.3	26.4	23	6.2	12.3	26.6	23.9	20.8	20.9	17.9
19.1	9.8	29.6	18	15.9	20.1	17.5	13.3	21.4	18.7
18.5	27	21.9	22.9	11.2	17	11	18.1	18	31.8

Sınıf: Değişkenin değer aralığı birbirinden kesin olarak ayrılmış gruplara bölünebilir, bu gruplara sınıf adı verilir. Frekans tablosu yapılmasında ilk problem sınıf sayısının belirlenmesidir. Sınıf sayısı k ile gösterilir. Sınıf sayısının 8-15 arasında olması istenir. Gözlem sayısı n olmak üzere sınıf aralıklarının sayısı yaklaşık olarak \sqrt{n} şeklinde hesaplanabilir. Sınıf sayısının belirlenmesinde aşağıda verilen Sturges formülü de kullanılabilir:

$$k = 1 + 3.3 \log(n)$$

Dağılım Sınırları: Örneklemdeki en küçük değer ile en büyük değere dağılım sınırları denir.

Dağılım Genişliği: Bir örneklemde en büyük değer ile en küçük değer arasındaki farka dağılım genişliği denir, R ile gösterilir.

Sınıf Aralığı: İki sınıf arasındaki farka sınıf aralığı denir, c ile gösterilir ve yaklaşık olarak

$$c = \frac{\text{dağılım genişliği}}{\text{sınıf sayısı}(k)}$$

şeklinde hesaplanır (c hesaplanan değerden daha küçük alınmamalıdır).

Alt Sınır: Bir sınıfın en küçük değeridir, A_s ile gösterilir.

Üst Sınır: Bir sınıfın en büyük değeridir, $Ü_s$ ile gösterilir.

Sınıf Orta Değeri: Bir sınıfın alt sınır ve üst sınır değerlerinin ortalaması sınıf değeridir, S_i ile gösterilir.

Frekans (Sıklık): Bir sınıfa düşen veri sayısına frekans denir, f ile gösterilir. Frekansların

toplamı veri sayısına eşit olmalıdır. Yani, $\sum_{i=1}^n f_i = n$ olmalıdır.

Görel Frekans (Frekans Yüzdesi): Her sınıfa düşen veri sayısının toplam veri sayısına

oranına denir, p ile gösterilir. $p_i = \frac{f_i}{n}$, $i = 1, 2, \dots, k$, $\sum_{i=1}^k p_i$

Tablo 1.1' de verilen 50 tane sülfür oksit miktarına ilişkin veriler için frekans tablosu oluşturulsun.

Sınıf sayısı olarak $k = 10$ alındı.

En büyük değer(max) = 31.8

En küçük değer(min) : 6.2

$R = 31.8 - 6.2 = 25.6$

(Sınıf aralığı) $c = \frac{R}{k} = \frac{25.6}{10} = 2.56 \cong 2.6$

Sınıflar $A_s \quad \dot{U}_s$	Sınıf orta değeri (S_i)	Frekans (f_i)	Eklemeli frekans	Görelî frekans f_i/n	Eklemeli Görelî frekans
[6.2, 8.8)	7.5	2	2	0.04	0.04
[8.8, 11.4)	10.1	4	6	0.08	0.12
[11.4, 14)	12.7	3	9	0.06	0.18
[14, 16.6)	15.3	6	15	0.12	0.3
[16.6, 19.2)	17.9	10	25	0.2	0.5
[19.2, 21.8)	20.5	8	33	0.16	0.66
[21.8, 24.4)	23.1	7	40	0.14	0.8
[24.4, 27)	25.7	5	45	0.1	0.9
[27, 29.6)	28.3	3	48	0.06	0.96
[29.6, 32.2)	30.9	2	50	0.04	1
		$\sum_{i=1}^k f_i = 50$		$\sum_{i=1}^k p_i = 1$	

Nitel Verilerde Frekans Tablosu

Örnek: Bir işyerinde çalışanların eğitim durumları araştırılmak istenmektedir. Bu amaçla rastgele 91 kişi seçilmiştir. Elde edilen sınıflar ;

Eğitim durumu	f_i	Görel frekans	Eklemeli f_i	Eklemeli Görel f_i/n
İlkokul	5	0.0549	5	0.0549
Ortaokul	12	0.1319	17	0.1868
Lise	22	0.2418	39	0.4286
Üniversite	39	0.4286	78	0.8521
Yüksek Lisans	9	0.099	87	0.956
Doktora	3	0.033	90	0.989
Doktora sonrası	1	0.0109	91	1

$$\sum_{i=1}^7 f_i = 91$$

$$\sum_{i=1}^7 p_i = 1$$

KAYNAKLAR

1. Uygulamalı İstatistik (1994)

Ayşen APAYDIN , Alaettin KUTSAL, Cemal ATAKAN

2. Olasılık ve İstatistik Problemler ve Çözümleri ile (2008)

Prof. Dr. Semra ERBAŞ

3. Olasılık ve İstatistik (2006)

Prof. Dr. Fikri Akdeniz

4. Olasılık ve İstatistiğe Giriş I-II (2011)

Prof. Dr. Fikri Öztürk

5. Fikri Öztürk web sitesi

<http://80.251.40.59/science.ankara.edu.tr/ozturk/index.html>