

İstatistik I

2. Hafta

21 Şubat 2022

2. Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

www.umitsarp.com

Ümit SARP, umit.sarp@ikcu.edu.tr

Giriş

- Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler
 - Veri kaynakları
 - Verilerin elde edilmesi
 - Anket Kavramı
 - Verilerin düzenlenmesi
 - Verilerin sınıflandırılması
 - Verilerin grafiklerle sunulması



Hatırlatma, 1. Hafta Konusu: Betimsel (tasviri) istatistik

İstatistik, geçmişini ve içinde bulunulan durumu tanımlayarak bir **veri kümesine** ilişkin **özet değerler** ve **grafikler** ortaya koyduğunda betimsel istatistik (descriptive statistics) adını alır. Özetleme ve verilerin yoğunlaştırılması diğer bir deyişle verilerin hacimleri azaltılarak kullanım değerlerinin artırılması betimsel istatistiğinin konusuna girer. **Tablolar ve grafikler** yardımı ile verilerin özetlenmesi ve çok sayıda sayıdan oluşan bir veri grubunun “ortalama” gibi tek bir sayıya indirgenmesi yine bu alan için geçerlidir [6].



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Kısaca betimsel istatistik bir veri kümesinde bulunan bilgiyi sayısal ve grafiksel yöntemleri kullanarak özetler ve sunar.

Bu kavramlara açıklayacak olursak;

Veriler, istatistik çalışmanın temelini oluşturmaktadır. İstatistiksel bir analizin doğruluğu ve güvenilirliği öncelikle verinin kalitesine (çalışmanın amacına uygun ve doğru bir şekilde toplanmış güvenilir verilere) bağlıdır [3].



Nicel Veri:

Bir deneyin sayılabilir, ölçülebilir özelliğini veren verilerdir. Sürekli nicel veri ve kesikli nicel veri olmak üzere iki türü vardır.

- **Sürekli nicel veri:** Ondalıklı değerler alabilen, ölçüm sonucu elde edilen verilerdir. Boy, kilo, ölçeklerden elde edilen puanlar gibi.
- **Kesikli nicel veri:** Sayılarak elde edilen ve tam sayılı değerlerdir. Hasta sayısı, öğrenci sayısı gibi [3].



Nitel Veri:

Bir deneğin niteliklerini belirten verilerdir.

- **Sınıflanabilen Nitel Veri:** Birbirinden bağımsız isim bildiren, kod ve numara ile gösterilebilen, sınıflara ayrılan verilerdir. Eğitim durumu, cinsiyet gibi.
- **Sıralanabilen Nitel Veri:** Belirli bir miktar belirtmeyen, bir sıra ya da dereceye göre elde edilen verilerdir. Örneğin, hastalığın evreleri gibi [3].



Veri Kaynakları

Verilerin elde edilme süreçleri çeşitlilik göstermektedir. Veriyi direkt olarak kaynağından sağlama imkanı var ise böyle elde edilen verilere **birincil veri** denir. Çeşitli kurum ve kuruluşların (DPT-Devlet Planlama Teşkilatı, SGK, TOBB, v.b.) yayınladıkları haftalık, aylık, yıllık, v.b. bültenlerden elde edilen veriler de birincil veri kapsamındadır [3].



Veri Kaynakları

Mesleki kuruluşlar (Mimarlar Odası, Makine Mühendisleri Odası, v.b.), medya kuruluşlarının ya da araştırma firmalarının yayınladıkları veriler ise daha ziyade **ikincil veri** kapsamındadır. Uluslar arası veri kaynaklarının başında ise OECD (Organization for Economic Cooperation and Develeopment), WB (World Bank), BM (UN: United Nations), IMF (International Money Fund) gelmektedir [3].



Verilerin Elde Edilmesi

İstatistiksel arařtırmaların yapılma nedenleri genellikle řu řekilde sıralanabilir:

- 1 Durum tespiti
- 2 Geleceęe dönük projeksiyon yapma

Durum tespitinden kasıt, bir kurumun, firmanın, haritanın, bilginin v.b. içinde bulunduęu durumun bilimsel olarak ortaya konmasıdır [3].



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Verinin toplanması denince ilk akla gelen ve en çok kullanılan yöntem **anket** yöntemidir. Anketlerin uygulanma şekilleri farklılık arz eder, örneğin anketörler vasıtasıyla yüzyüze görüşme yaparak olabileceği gibi, e-posta, posta v.b. yöntemlerle de anket formları göndermek suretiyle veri elde edilebilir. Ancak posta yoluyla yapılan anketlere büyük olasılıkla geri dönüş yapılmamaktadır veya en iyi ihtimalle yarım cevaplanmış, bir sürü cevapsız soru bırakılmış anket formlarıyla karşılaşmaktadır [3].



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Bu nedenle en sağlam veri toplama şekli yüzyüze görüşme yaparak anket yapmaktır denebilir. Gelişmiş ülkelerde posta yoluyla anket cevpalama oranı oldukça yüksektir, her ne konuda olursa olsun bir anket muhakkak bir iyileştirme çabasına hizmet etmektedir, böyle düşünüldüğünde, anketi cevaplamanın ne kadar önemli olduğu anlaşılabilir. Ülkemizde ise bırakın postayla anket cevaplamayı, bazılarının, beş dakikalığına yolunu çeviren anketörleri bile azarladıklarını görmekteyiz [3].



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Halbuki artık siyasi partiler bile araştırma şirketlerine anketler yaptırarak halkın nabzını tutmakta, neredeyse seçim sonuçlarıyla aynı denecek yakınlıkta tespitler yapmaktadırlar. Bu sayede politik planlamalar daha gerçekçi yapılabilmekte, kaynaklar daha etkin (gereksiz israf olmadan) kullanılabilir [3].



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

O halde işini zamanında doğru düzgün yapmaya çalışan, halkın fikirlerine değer veren ve kaynaklarını etkin bir biçimde kullanmak isteyen herkese bunu bir yardım talebi gibi kabul ederek yardım etmeliyiz. Bakış açımız gerçekten de böyle olmalıdır, bu sadece basit bir örnektir ama binlercesi düşünülebilir.

Anketin öneminden bahsettikten sonra, anketin nasıl uygulanması gerektiğinden bahsedelim. Adım adım özetlersek;



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

- 1 Anket konusunun ve amacının belirlenmesi
- 2 Anketin uygulanacağı birimlerin belirlenmesi (genel olarak herkes mi, yoksa sadece bayanlar mı, ya da sadece 20-25 yaş arası gençler mi, v.b.)
- 3 Örneklemeye yöntemine karar verilir. (bu konu örneklemenin konusudur ve çeşitli yöntemler vardır, örneğin basit rastgele örneklemeye, sistematik örneklemeye, tabakalı örneklemeye, küme örneklemesi, v.b.)
- 4 Çerçevenin belirlenmesi (çerçeve anket uygulanacak birimlerin listesi gibi düşünülebilir, böylece çalışmanın sınırları çizilmiş olur, örneğin anket bir ilçede uygulanacaksa her mahalleden kaç kişiye anket uygulanacağını belirlemek çerçeveyi çizmek anlamına gelir)



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

- 5 Anket yüzyüze görüşme yöntemi ile uygulanacaksa, bu işi en iyi ve doğru şekilde yapabilecek, konuyla ilgili bilgilendirilmiş anketörlerin seçilmesi. (örneğin inşaat malzemeleri üreten bir firma, yapı malzemelerinin tercihi konusunda anket yaptırmak istediğinde, malzemelerin özelliklerini ve ne işe yaradığını bilmeyen konudan bihaber anketörlere görev vermek sizce ne kadar doğru olur?)



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

- 6 Anketin sahada uygulanması ve verilerin toplanması. (anketin uygulanması başlı başına uzmanlık isteyen bir durumdur. Öyle ki, anketörün tutum ve tavrı, soru sorma stili, ses tonu, giydiği kıyafete kadar özenle tespit edilmelidir. Örneğin bir diş macunu firmasının anketinde, 32 diş pırıl pırıl parlayan çok güzel dişlere sahip bir anketöre bu işi yaptırırsanız, belki haftada bir diş fırçalayan biri, günde üç kez diş fırçalamakta olduğunu beyan edebilir, çünkü anketörün görünümünü karşısında ezilmiş ve kendini kötü hissettiği için de maalesef yanlış beyanda bulunmuştur.)



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Verilerin anket yoluyla toplanması genel olarak bu biçimdedir.
Örnekleme konusu ilerleyen haftalarda ele alınacaktır.



Verilerin Düzenlenmesi

Aşağıdaki örnek veriyi kullanarak verilerin düzenlenmesi ve özetlenmesi konularını ele alacağız:



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Örnek Veri:

19	13	5	4	1	9	9	7	11	11
12	1	1	17	13	18	5	2	12	14
16	18	20	19	19	14	17	15	2	3
6	3	12	16	13	13	11	10	10	8
10	4	2	5	15	10	14	10	1	7

Örnek veri seti, 1 ile 20 arasında değer alan 50 adet gözlemden oluşmaktadır.



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Dizi:

Bir veri setinin dizi haline getirilmesi demek, ya küçükten büyüğe, ya da büyükten küçüğe sıralamaktır. Genellikle küçükten büyüğe sıralama yapılır. Diziye bakarak en küçük değer ile en büyük değer arasındaki farkı belirleyerek, böylece gözlemlerin hangi aralıkta dağıldıklarını görebiliriz. Dizideki en küçük değere X_{min} , en büyük değere X_{max} , aralarındaki farka ise “açıklık” yani “range”, “değişim aralığı” da denebilir.



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Örnek Dizi:

1	1	2	2	3	3	4	4	5	5
5	6	6	6	7	7	7	7	8	8
8	9	9	9	9	10	10	10	10	10
10	11	11	11	11	12	12	13	13	13
14	14	15	16	16	17	18	18	19	20

$X_{min} = 1$, $X_{max} = 20$, $range = 20 - 1 = 19$



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler



Figure: [kaynak: figy.com.tr]



Frekans:

Veri kümesine bakarak verideki her bir değerin kaç defa tekrar ettiğini görmek mümkündür. Grafikselleştirme yoluyla başvurmaksızın, her bir gözlem değerini gösteren tabloya **frekans** (sıklık) tablosu denir. f_i frekans (sıklık) dağılımını elde etmiş oluruz.

Not: Böyle serilerin grafik sunumunda çizgi diyagramı kullanılır.



Örnek Frekans Tablosu:

Örnek veri setimizin frekans dağılımını oluşturalım:

X_i	f_i	X_i	f_i
1	2	11	4
2	2	12	2
3	2	13	3
4	2	14	2
5	3	15	1
6	3	16	2
7	4	17	1
8	3	18	2
9	4	19	1
10	6	20	1

Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

$$\text{Toplam } \sum_{i=1}^{20} f_i = 50$$

Frekansların toplamı; $\sum_{i=1}^k f_i$, gözlem sayısı'na eşittir. Burada k , sınıf sayısıdır.

Not: Frekans dağılımının grafiksel gösteriminde çoğunlukla çizgi diyagramına başvurulur.



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Gözlem sayısı çoğaldığında, her bir verinin çizgi diyagramında olduğu gibi tek tek gösterilmesi güçleşmektedir. Bu nedenle, büyük veri seti söz konusu olduğunda verilerin sınıflandırılmaları yoluna gidilir. Verinin belli sınıflara bölünerek bu aralıklara karşılık gelen frekanslarıyla birlikte gösterildiği tabloya, sınıflandırılmış frekans dağılımı denir. Sınıflandırılmış frekans dağılımını oluşturmak için izlenmesi gereken adımlar şöyledir:



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Sınıflandırılmış Frekans Verisi:

1. Verinin değişim aralığı (açıklığı- range'i) hesaplanır.
2. Değişim aralığı eşit uzunlukta alt aralıklara yani sınıflara bölünür. Sınıfların üst sınırı ile alt sınırı arasındaki fark o sınıfın sınıf genişliğini verir. Sınıf orta noktası, sınıfın alt sınırı ile üst sınırının ortalamasıdır. Sınıf sayısını bulmak için aşağıdaki formüllerden herhangi biri kullanılabilir;



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

k , sınıf sayısı ve n , veri setindeki gözlem sayısını ifade etmek üzere;

$$k = \sqrt{n}$$

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

$$k = 5 \log n$$

Değişim aralığı, sınıf sayısı ve sınıf genişliği arasındaki ilişki şöyledir;

$$\text{sınıf genişliği} = \frac{\text{range}}{k}$$

- Her bir sınıf aralığına kaç gözlemin düştüğü sayılır. Bu sayılar o sınıfın frekansını verir.



Örnek Sınıflandırılmış Frekans Verisi:

Örnek veri setimizi sınıflandırılmış frekans serisi haline getirelim. İlk olarak verinin değişim aralığını buluyoruz:

$$X_{max} = 20,$$

$$X_{min} = 1,$$

$$Range = 19$$

Sonra, k sınıf sayısını buluyoruz:

$k = 1 + 3.3 \log(50) = 6.6$ bulduk, fakat k tamsayı olmak zorundadır, o halde $k = 7$ alalım. Sınıf genişliği = $19/7 = 2.714$ bunu da 3'e yuvarlayalım.

İlk sınıfı bulmak için, X_{min} değerine sınıf genişliğini ekleyelim. Böylece ilk sınıfın üst sınırını da bulmuş olacağız.

Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Örnek Sınıflandırılmış Frekans Verisi:

Sınıflar	Frekanslar
1 – 4'den az	6
4 – 7'den az	8
7 – 10'den az	11
10 – 13'den az	12
13 – 16'den az	6
16 – 19'den az	5
19 – 22'den az	2



Basit Serinin Sunumu İçin Nokta Diyagramı:

Verinin değişim aralığını göstermek üzere ölçeklendirilmiş bir doğru üzerinde her bir gözlemin bir nokta ile temsil edildiği bir grafikdir.



Örneğimizin Nokta Diyagramı:



Figure: Basit Serinin Sunumu İçin Nokta Diyagramı, [3]



Frekans Serisinin Sunumu İçin Çizgi Diyagramı:

Veri kümesine bakarak verideki her bir değerin kaç defa tekrar ettiğini görmek mümkündür. Grafikselle bir yola başvurmaksızın, her bir gözlem değerini frekansıyla (sıklığıyla) birlikte gösteren bir tablo düzenlersek, frekans (sıklık) dağılımını elde etmiş oluruz. Böyle serilerin grafik sunumunda çizgi diyagramı kullanılır.



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Örneğimizin Çizgi Diyagramı:

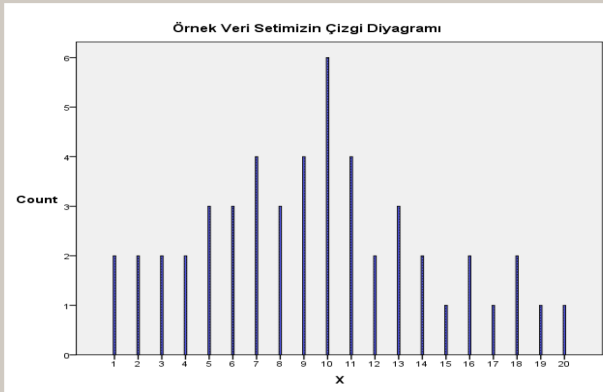


Figure: Frekans Serisinin Sunumu İçin Çizgi Diyagramı, [3]

Histogram:

Histogram, dikey ekseninde sıklıkların, yatay ekseninde sınıf aralıklarının bulunduğu, yan yana dikdörtgen kutucuklardan oluşan bir grafik türüdür. Her bir kutu bir sınıfa tekabül eder ve kutunun alanı, ilgili sınıfın sıklığını verir.



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Kutunun alanı=taban \times yükseklik

Burada taban, sınıf genişliğidir, yani ilgili sınıfın üst sınırı ile alt sınırı arasındaki farktır.

Bu değeri “ h ” ile ifade edersek, yükseklik şuna eşittir:

Kutunun yüksekliği= $\frac{f_i}{h}$ 'tır.

Bu durumda kutunun alanı= $h \times \frac{f_i}{h} = f_i$ 'dir yani ilgili sınıfın sıklığı (ilgili sınıftaki birim sayısı) dir.



Örneğimizin histogramı:

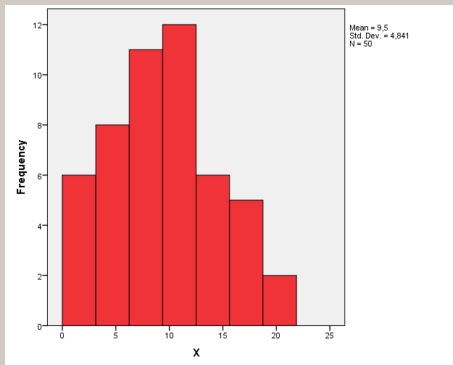


Figure: Histogram, [3]

Box-Whisker Grafiği:

Bu grafik türünde beş adet istatistiksel özet bir arada görülür. Bunlar;

X_{\max} ,

X_{\min} ,

Q1 (1. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %25'ten kesen değer),

Q2 (2. Çeyrek, asıl adı medyandır, küçükten büyüğe sıralanmış bir seriyi %50'den kesen değer),

Q3 (3. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %75'ten kesen değer).



Box-Whisker Grafiği:

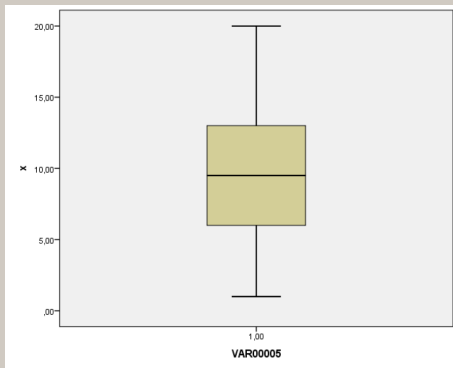


Figure: Box-Whisker Grafiği, [3]

Daire Grafiđi:

Daire řeklindedir. Bu grafiđi çizmek için her bir sınıfın sıklıđının toplam sıklık içindeki oranı bilinmelidir. Toplam sıklık 360^0 olmak üzere, sınıflar sıklıkları oranında birer dilimle temsil edilirler.



Daire Grafiđi:

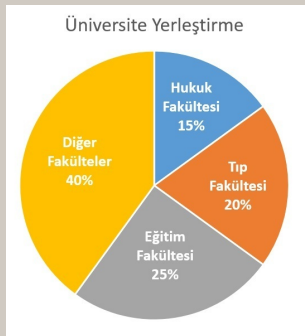


Figure: Daire Grafiđi



Veri Türleri, Toplama Yöntemleri, Sınıflama ve Grafikler

Ödev:

Bir kazı ekibinde 100 kişi vardır. Her birinin bulduğu kıymetli eserlerin sayısı aşağıda verilmiştir. Buna göre, X_{max} , X_{min} , range, sıralı dizi tablosu, frekans tablosu ve sınıflandırılmış frekans tablosunu bulunuz.

1	91	29	1	74	1	48	73	34	83
75	20	55	25	27	41	82	98	8	34
36	38	46	52	3	78	51	57	94	56
81	89	89	83	88	12	15	47	82	95
49	93	3	89	21	47	24	43	18	74
76	85	87	54	70	66	85	70	35	16
30	72	78	47	92	84	95	41	64	82
1	84	63	51	35	70	20	25	94	18
80	4	62	98	43	21	42	9	71	89
41	81	2	40	19	18	22	88	56	39



3. Hafta: Merkezi Eğilim Ölçüleri

İpucu : Aritmetik Ortalama, Mod, Medyan,...



Kaynaklar I

- [1] K. Mert Çubukçu,
"Planlamada ve Coğrafyada Temel İstatistik ve Mekansal İstatistik",
Ankara: Nobel Akademik Yayıncılık Eği. Dan. Tic. Ltd. Şti., (2015).
- [2] A. Özmen, F. Er, M. Atlas, E. Şıklar,
"İstatistik (AÖF)",
Eskişehir: Anadolu Üniversitesi Yayınları, (2012).
- [3] L. İşbilen Yücel,
"İstatistik Maliye Uzaktan Eğitim",
İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi Ders Notu.
- [4] Ö. Serper,
"Uygulamalı İstatistik 1",
Bursa: Ezgi Kitapevi, (2004).



Kaynaklar II

- [5] Murat Komisyon,
"İstatistik",
Murat Açıköğretim Yayınları, (2004).
- [6] N. Gürsakal, A. Oğuzlar,
"Betimsel İstatistik",
Dora Yayıncılık, (2019).
- [7] Y. Baykul, C. O. Güzeller,
"Sosyal Bilimler için İstatistik Uygulamaları",
Ankara: Pegem Akademi, (2014).

