



Olasılık ve Raslantı Değişkenleri

Yrd.Doç.Dr. Halil YİĞİT
Bilişim Sistemleri Mühendisliği

[Ders Kitabı ve Değerlendirme]

■ Ders Kitabı

- Olasılık ve İstatistik / Fikri AKDENİZ
- İstatistik formülleri ve olasılık dağılımları / Fikret Er, Harun Sönmez
- Probability&Statistics for Engineers&Scientists / Ronald E. Walpole and et al.

■ Sınav

- Vize (%40)
- Final (%60)

[İstatistik Nedir?]

■ Tanım:

- Verilerin toplanması, düzenlenmesi, uygun yöntemlerle analiz edilmesi ve verilerden geçerli bir sonuç çıkarılmasına istatistik denir.

■ Uygulama Alanları

- Nüfus sayımında
- Hava tahminlerinde
- TV reytinglerin belirlenmesinde
- Seçim tahminlerinde
- Risk analizinde
- Ve daha bir çok alanda.....

İstatistiğin Konusu Olan Olaylar

- Olaylar ikiye ayrılmaktadır:
 - Tipik Olay
 - Kollektif (Toplu) Olay
- Tipik Olaylar
 - Birbirinin tam benzeri olaylar olup, gerekli koşullar oluştuğunda her zaman aynı şekilde tekrar eden ve az sayıda faktör tarafından etkilenen olaylardır.
 - Fiziksel ve kimyasal olaylar tipik olay olarak örnek verilebilirler.
 - İnsanların 2 gözlü doğması, hidrojen ve oksijenin belirli koşullarda suyu meydana getirmesi,..

İstatistiğin Konusu Olan Olaylar

■ Kollektif (Toplu) Olaylar

- Birbirine benzemeyen, ortak yönleri olmasına karşın aralarında farklılıklar bulunan olaylardır.
- Birçok faktör tarafından belirlenir. Dolayısı ile birkaç deneme ile olayların genel karakterleri belirlenemez.
- Tüm faktörler belirlenebilse bile hepsi kontrol edilemez
- Nüfus kollektif olay için iyi bir örnektir. Nüfusu oluşturan bireylerin, insan olmak ve aynı bölgede veya ülkede yaşamak gibi ortak özellikleri olmasına rağmen cinsiyet, yaş, meslek gibi çeşitli özellikler bakımından farklıdırlar. Bu farklılıklar nedeniyle bütün nüfusu temsil edebilecek tek bir insan bulmak mümkün olmaz.
- Bir ağacın yaprakları ve meyveleri arasında da şekil, renk, büyüklük, ağırlık, sertlik ve su oranı bakımından farklılıklar vardır.

[İstatistiğin Konusu Olan Olaylar]

- Bütün bu açıklamalardan sonra diyebiliriz ki;
 - İstatistik tipik olaylarla ilgilenmez.
 - İstatistik toplu ve sonuçları sayısal olarak ifade edilebilecek olaylar ile ilgilendiği için **Kollektif Olaylar** istatistiğin konusunu oluşturur.

[Temel Kavramlar]

■ Ana Kütle:

- Kollektif olay özelliğinde ve aynı cinsten (homojen) birimlerin meydana getirdiği topluluktur. Birimler tamamen aynı özelliklere sahip olmasalar da, bazı ortak yanlarının bulunması gereklidir.
- Örneğin **yıl** bir kütle olarak alınırsa günler birimdir.
- Birimleri sayılabilen kütlelere **belirli kütle**, sayılamayanlara **belirsiz kütle** adı verilir.
- Bir ülkenin nüfusu, bir şehirdeki binalar belirli kütle, bir nehirdeki balıklar, ormandaki karıncalar sayılamayacağı için belirsiz kütledir.
- **Kütleler sürekli veya süreksiz olarak da sınıflanabilirler**. Arsa, tarla gibi birbirine bitişik olan birimler sürekli, insan, otomobil gibi birimler süreksiz kütleleri oluştururlar.

[Temel Kavramlar]

■ Birim:

- Kütleyi oluşturan kollektif olayların her birine birim adı verilir.
- Canlılar (insan,hayvan) sosyal bir kuruluş (aile, şirket) bir olay (doğum,ölüm,evlenme) birim olarak kabul edilir.
- Birimler mutlaka sayılabilir veya ölçülebilir özelliklere sahip olmalıdırlar.
- Birimler aynı zamanda homojen olmalıdırlar.

■ Örnek:

- Ana kütleden seçilen birimlerin oluşturduğu alt topluluk örnek olarak adlandırılır.
- Diğer bir deyişle inceleme yapılan ana kütlenin bir bölümüdür.
- Bir firmada satın alınan hammaddenin tamamı ana kütleyi, kalite kontrolü için alınan parça örneği oluşturur.

[Temel Kavramlar]

■ Veri:

- **Tanım 1:** Bir sonuca varabilmek için gerekli olan ilk bilgi ya da anlam çıkartmada, sonuca varmakta kullanılan nicelikler, olaylar, kanıtlar, sayı kümeleri veya semboller olarak tanımlanır.
- **Tanım 2:** Henüz işlenmemiş kanıtlardır.
- **Tanım 3:** Bir araştırmacı tarafından gözlemlerden elde edilen sayısal olan ya da olmayan sonuçlardır.
- **Örneğin;** Bir nesnenin rengi, ağırlığı; bireyin belli bir konudaki görüşleri tutum ve davranışları birer veridir

[Temel Kavramlar]

■ Değişken:

- Birimlerin sahip olduğu özelliklere değişken denir.
- Değişebilen yani en az iki değer alabilen her şey değişkendir.
- Bir başka deyişle gözlemden gözleme değişik değerler alabilen objelere, özelliklere ya da durumlara değişken denir.
- Değişen şeyi ifade etmek için kullanılan bir kavramdır.
- **Örneğin;** cinsiyet (k-e), yaş (1,2,3,.....), iyileştirme şekilleri (ilaç,psikolojik yardım) birer değişken olarak alınabilir.

[Temel Kavramlar]

■ Değişken Türleri:

○ Nicel (Süreli) Değişken

- Alt ve üst sınırları arasında herhangi bir değer alabilme olasılığına sahip değişkenler, bir başka deyişle kesirli olarak ifade edilebilen değişkenler sürekli değişkenlerdir.
- Bunlar matematiksel olarak yani sayısal ölçekte ifade edilebilir. Miktar hakkında bilgi verir.
- Zeka, yaş, ağırlık, hava sıcaklığı, hız, nüfus, hesap bakiyesi, pil ömrü, öğrenci sayısı vb.

[Temel Kavramlar]

■ Değişken Türleri:

○ Nitel (Süreksiz) Değişken:

- Sayısal olmayan, daha çok sözcük ve sembollerle ifade edilen ve gözlemden gözleme kalite yönünden değişiklik gösteren değişkenlerdir.
- Cinsiyet, din, milliyet, medeni durum , göz rengi vb.
- Örneğin medeni durum için bekar, sözlü, nişanlı, evli ve dul olmak üzere beş değer verilebilir. Normal olarak bunlar arasında altıncı bir değer verilemez.

[Temel Kavramlar]

■ Tam Sayım

- Ana kütle ile ilgili bilgi toplamak istendiğinde tüm birimler teker teker inceleniyor ise bu işleme tam sayım denir
- Tam sayımda ana kütlelerin özellikleri kesin olarak öğrenilebilir.
- Dezavantajları; Zaman Alıcıdır, Masraflıdır (Nüfus Sayımı)

■ Örnekleme

- Ana kütlede ana kütle birim sayısından daha az sayıda rastgele birimin seçilerek; bu birimler yardımıyla ana kütle parametrelerinin tahmin edilmesi işlemidir.
- Toplanan örneklerin incelenen ana kütleyle tam anlamıyla yansıtması gerekmektedir.

[Temel Kavramlar]

■ Örneklemenin Üstünlükleri

- Kısa sürede ve daha az masraf yapılabilir
- Veri toplama ve analizde yapılabilecek hatalar daha azdır.

■ Örneklemenin Sakıncaları

- Parametre tahminlerinde sapmalar ortaya çıkar.

Verilerin Düzenlenmesi

- **Veri toplama yöntemleri:**

- Anket
- Gözlem
- Deney

- Veriler hangi yöntemle toplanırsa toplansın, elde edilen veriler genellikle istatistiksel analize hazır değildir.

- Araştırmalarda elde edilen veriler genellikle düzensiz ham verilerdir.

- Bu verilerin analize uygun hale getirilmesi için kolay ve anlaşılır bir biçimde düzenlenmeleri gerekir.

- Bu düzenleme çeşitli şekillerde yapılabilir:

- Basit Seriler
- Frekanslı Seriler
- Gruplandırılmış Seriler

[Basit Seriler]

- Elde edilecek ham verilerin küçükten büyüğe doğru sıralanması ile elde edilen serilere **basit seri** denir.
- Basit bir seri birden fazla birimden oluşur ve birim sayısı n ile gösterilir
- Serinin i 'ninci elemanı X_i değişkeni ile gösterilir.
- Basit serinin toplam değeri X_i 'lerin toplamına eşit olacağından,

$$\text{Serinin toplam değeri} = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

- **Örnek:** Bir işletmede 25 işçiye verilecek çocuk paraları ile ilgili bir araştırma yapılmaktadır. İşçilerin çocuk sayıları aşağıda verilmiştir.

1,3,2,2,3,1,4,5,3,6,0,5,2,3,2,4,8,0,1,2,3,3,1,0,4

- Verilen değerleri basit seri şeklinde düzenleyelim:

0,0,0,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,5,5,6,8

[Frekans Seriler]

- İncelenecek birim sayısı arttıkça basit seriler çok uzun olacaktır.
- Bu durumda çalışma kolaylığı sağladığı için frekans serileri kullanılır.
- Frekans serilerinin değişkenin çok sayıda farklı değer almadığı durumlarda kullanılması daha uygun olur.
- Düzenlenen bir frekans serisi iki sütundan oluşur:
 - **Birinci Sütun:** Değişkenin aldığı farklı değerler yer alır
 - **İkinci Sütun:** f_i ile gösterilir ve değişkenin aldığı değerlerin tekrar sayısı gösterilir.

X_i	f_i
X_1	f_1
X_2	f_2
X_3	f_3
.	.
.	.
.	.
X_k	f_k

Değişkenin k sayıda farklı değer aldığı bir frekans serisi

[Frekans Seriler]

- Frekans serisinin toplam değerini aşağıdaki gibi hesaplayabiliriz:

$$f_1X_1 + f_2X_2 + \dots + f_kX_k = \sum_{i=1}^k f_iX_i$$

- **Örnek:** Daha önce basit seri olarak düzenlenen seri frekans serisi olarak düzenlenirse:

X_i	f_i
0	3
1	4
2	5
3	6
4	3
5	2
6	1
8	1
	<hr/>
	25

[Gruplandırılmış (Sınıflı) Seriler]

- İncelenecek birim sayısının çok fazla olması durumunda frekans serileri de uzun sayı dizilerine dönüşür
- Bu durumda gruplandırılmış seriler düzenlenir
- Gruplandırılmış seriler iki sütundan oluşur:
 - **Birinci Sütun:** Sınıflar sütunudur
 - **İkinci Sütun:** Frekans sütunudur
- *Gruplandırılmış serilerde en önemli nokta sınıf sayısının kaç tane veya sınıf aralığının ne olacağının belirlenmesidir.*
- Sınıf aralığı belirlendikten sonra
 - İlk sınıfın alt sınır değeri sınıflar sütununa yazılır
 - Alt sınır değere sınıf aralığı eklenerek üst sınır değeri elde edilir
 - Her sınıfın üst sınıf değeri bir sonraki sınıfın alt sınır değerini oluşturur

[Gruplandırılmış (Sınıflı) Seriler]

- **Örnek:** Daha önce frekans serisi olarak düzenlenen örneği 2 eşit aralıklı sınıflı seri olarak düzenleyiniz.

Frekans Serisi

X_i	f_i
0	3
1	4
2	5
3	6
4	3
5	2
6	1
8	1
	<hr/> 25

Sınıflandırılmış Seri

Sınıflar	f_i
0-2	7
2-4	11
4-6	5
6-8	1
8-10	1

*** Sınıflardaki frekanslar belirlenirken alt Sınıf değeri dahil üst sınıf değeri **hariç** tutulur

$$\text{Yaklaşık sınıf aralığı} = \frac{(\text{en büyük değer} - \text{en küçük değer})}{\text{sınıf sayısı}}$$

[Gruplandırılmış (Sınıflı) Seriler]

- Gruplandırılmış seriler ile işlem yapılırken frekansların her bir sınıf içerisinde eşit biçimde dağıldığı düşünülür.
- Bu nedenle sınıf ortaları X_i olarak düşünülerek serinin toplamı frekans serilerinde olduğu gibi hesaplanır

$$\text{Serinin toplam değeri} = f_1m_1 + f_2m_2 + \dots + f_k m_k = \sum_{i=1}^k f_i m_i$$

$$m_i = \frac{X_{\min}^i + X_{\max}^i}{2} \text{ i'ninci sınıf ortası } m_i, \text{ sınıf alt sınır ve üst sınır toplamının ikiye bölünmesi ile bulunur}$$

- **Örnek:** Bir önceki örnekteki gruplandırılmış serinin toplamını hesaplayınız

Sınıflar	f_i	m_i	$f_i m_i$
0-2	7	$(0+2)/2=1$	7
2-4	11	$(2+4)/2=3$	33
4-6	5	$(4+6)/2=5$	25
6-8	1	$(6+8)/2=7$	7
8-10	1	$(8+10)/2=9$	9
			$\sum_{i=1}^k f_i m_i = 81$

[Gruplandırılmış (Sınıflı) Seriler]

- Gruplandırılmış serilerde sınıflardaki verilerin sınıf aralığında düzgün dağıldığının varsayılması hesaplanan ölçülerde sapmalara neden olacaktır:

- Basit seride,
$$\sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$$
$$= 0 + 0 + 0 + \dots + 5 + 6 + 8 = 68$$

- Frekans serisinde,

$$\sum_{i=1}^n f_i X_i = f_1 X_1 + f_2 X_2 + \dots + f_k X_k$$
$$= 0(3) + 1(4) + 2(5) + \dots + 6(1) + 8(1) = 68$$

Birikimli Seriler

- Bir frekans dağılımında, her sınıfın frekansına bir önceki sınıfın frekansı eklenerek oluşturulan seriye “**birikimli seri**”, bu tür oluşturulan frekanslara da “**birikimli frekanslar**” adı verilir.
- Birikimli seriler, küçükten büyüğe ya da büyükten küçüğe doğru oluşturulabilirler.
- Eğer birikimli seriler küçükten büyüğe doğru oluşturulmuşsa “**-den az**”, büyükten küçüğe doğru oluşturulmuşsa “**-den çok**” olarak isimlendirilirler

Ağırlık Sınıfları (kg)	Frekanslar f	(-den az)	(-den çok)
1.50 - 1.75	5	5	$95 + 5 = 100$
1.75 - 2.00	6	$6 + 5 = 11$	$85 + 6 = 95$
2.00 - 2.25	10	$10 + 11 = 21$	$79 + 10 = 89$
2.25 - 2.50	10	$10 + 21 = 31$	$69 + 10 = 79$
2.50 - 2.75	35	$35 + 31 = 66$	$34 + 35 = 69$
2.75 - 3.00	15	$15 + 66 = 81$	$19 + 15 = 34$
3.00 - 3.25	13	$13 + 81 = 94$	$6 + 13 = 19$
3.25 - 3.50	2	$2 + 94 = 96$	$4 + 2 = 6$
3.50 - 3.75	4	$4 + 96 = 100$	4
	100		

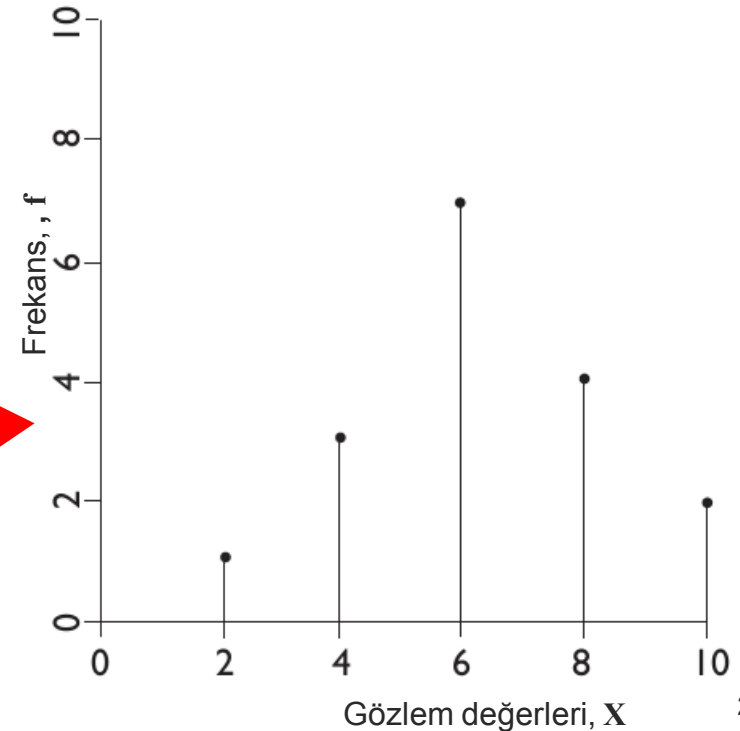
Bir doğum evinde doğan 100 bebeğe ilişkin sınıslandırılmış serinin küçükten büyüğe ve büyükten küçüğe doğru birikimli serileri

[Serilerin Grafikle Gösterilmesi]

- Frekans serilerinde **frekanslar gözlem değerlerine göre değiştiğinden** gözlem değerleri yatay eksen, frekanslar ise dikey eksen üzerinde gösterilir. Grafik, yatay eksen üzerinde belirlenen değerlerden uzunlukları ilgili frekanslar kadar olan dikey doğru parçalarıyla oluşturulur. Bu tür grafiklere “**çubuk grafik**” adı verilir

X	f
2	1
4	3
6	7
8	4
10	2
	17

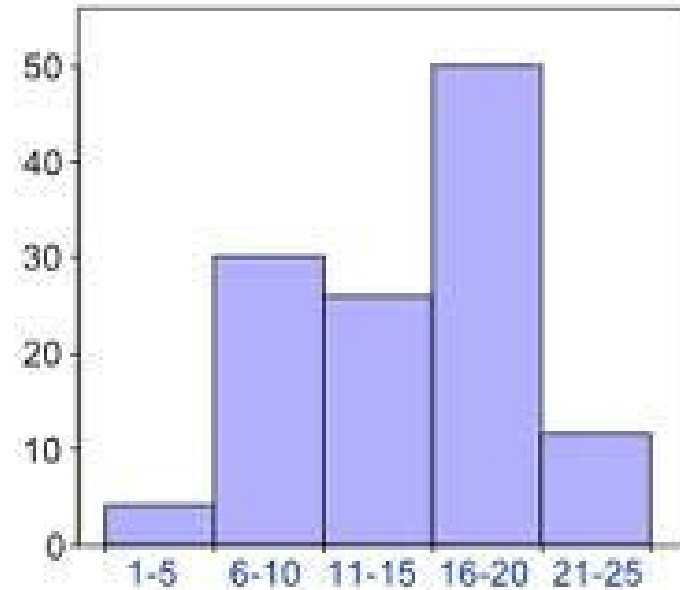
Verilen serinin çubuk grafiği



Sınıflandırılmış Serilerin Grafikle Gösterilmesi

■ Histogram

- Alanı ilgili sınıfın frekansına ve tabanı da ilgili sınıfın aralığına eşit, birbirine bitişik dikdörtgenlerden oluşan bir grafik gösterimdir.
- Bir histogram çizilmeden önce, sözü edilen dikdörtgenlerin uzunluklarının ayarlanması gerekir. Bunun için **frekanslar sınıf aralığına bölünerek**, dikdörtgenlerin alanları ilgili sınıfların frekanslarına eşit hale getirilir.



Örnek Histogram

Histogram

Sınıflar	frekanslar
0 – 4	12
4 – 8	16
8 – 12	20
12 – 16	24
16 – 20	20
20 – 24	8
	100

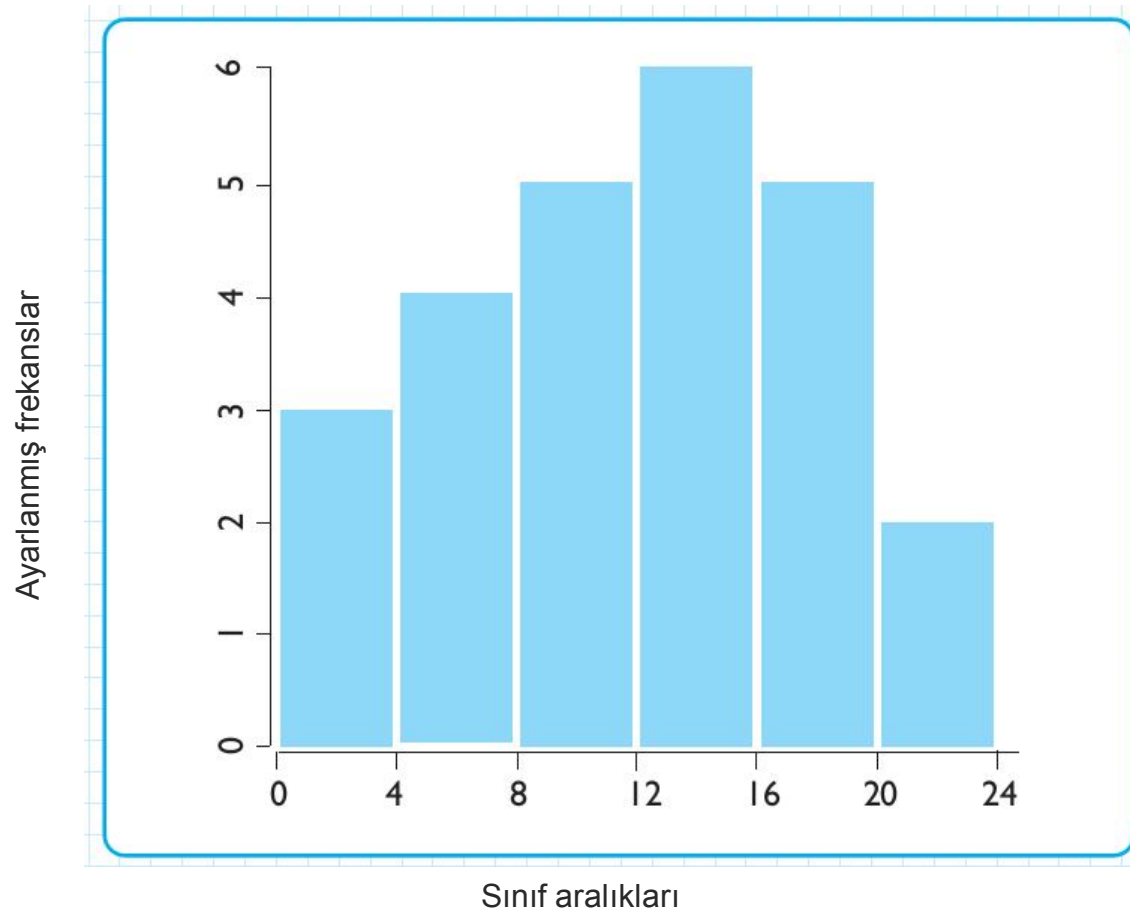
Sürekli bir X değişkenine ilişkin gözlem sonuçları

Sınıflar	f	Sınıf Aralıkları h	Ayarlanmış Frekanslar f / h
0 – 4	12	4	$12 / 4 = 3.0$
4 – 8	16	4	$16 / 4 = 4.0$
8 – 12	20	4	$20 / 4 = 5.0$
12 – 16	24	4	$24 / 4 = 6.0$
16 – 20	20	4	$20 / 4 = 5.0$
20 – 24	8	4	$8 / 4 = 2.0$
	100		

Ayarlanmış frekansların elde edilmesi

[Histogram]

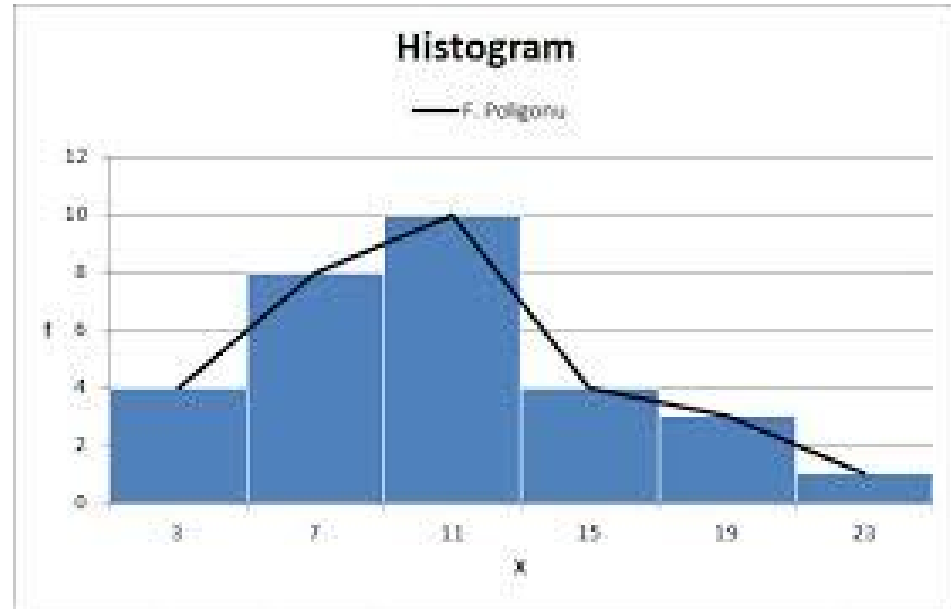
Birinci ve son sütundan yararlanılarak histogram çizilir:



*** Eğer verilen seride sınıf aralıkları eşit değilse, histogram yine aynı yöntemle oluşturulur.

[Poligon (Frekans Çokgeni)]

- Histogramın tepe orta noktalarının birleştirilmesiyle elde edilen, sınıflandırılmış serilere ilişkin, diğer bir grafik türüdür.
- Histogramın tepe orta noktaları, ilgili sınıflara ilişkin değişkenlerin değerlerini ifade ettiğinden, frekans poligonu, değişkenlerin değerlerine göre oluşturulmuş bir grafiktir.



Poligon (Frekans Çokgeni)

Sınıflar	frekanslar
10 – 15	10
15 – 20	15
20 – 25	20
25 – 30	30
30 – 35	15
35 – 40	10
40 – 45	5
	105

Sınıflar	f	X	h	f / h
10 – 15	10	12.5	5	2
15 – 20	15	17.5	5	3
20 – 25	20	22.5	5	4
25 – 30	30	27.5	5	6
30 – 35	15	32.5	5	3
35 – 40	10	37.5	5	2
40 – 45	5	42.5	5	1
	105			

