

1. BÖLÜM

İSTATİSTİĞE GİRİŞ

İSTATİSTİK KAVRAMLARI

1.1. Tanım: Herhangi bir konuyu araştırmak amacıyla belirlenen kurallarla toplanan bir takım veri (nümerik veya anti nümerik) bilgileri toplanmasıyla, toplanan bu sayısal bilgileri değerlendirerek büyük kümeler hakkında bilgiler elde etmeye çalışan bilime istatistik denir. Elde edilen alt kümlerde bilgilerin işlenmesi, değerlendirilmesi, yorumlanması, genellemesi ve bazen de tahmin yapma konularında İstatistik çalışmalar yapar.

İstatistik, sağlık, işletme, iktisat, eğitim, mühendislik, turizm, ziraat gibi birçok alanda bazı bilgilerin belirlenmesinde kullanılır. İstatiksel hesaplar;

- i) Tablolar ve grafikler
- ii) Çeşitli sayısal veriler

gibi sonuçlarla işlem yapılmaya çalışılır.

İstatistik, araştırmaların amacına yönelik olarak var olan bir problemin belirlenmesi, bunların ilişkin değişkenlerin oluşturulması, değişkenlere ilişkin verilerinin toplanması, bu verilerden tablolar ve sayısal değerler elde ederek çözümlenmesi ve yorumlanması sürecini içeren yöntemler bilimidir.

Örnek: Bir apartmanda bulunan 8 aile üzerinde yapılan ankette;

1. soru: "Evinizde internet var mı?"
2. soru: "Evinizdeki televizyon HD, LCD türünden mi?"
3. soru: "Evinizde kaç tane Android cep telefonu var?"

şeklinde. Anket sorularına elde edilen serilerden cevapları, bir tablo halinde şöyle düzenleyebiliriz.

	1. Daire	2. Daire	3. Daire	4. Daire	5. Daire	6. Daire	7. Daire	8. Daire
1. Soru	Var	Yok	Var	Var	Yok	Yok	Var	Var
2. Soru	Var	Var	Var	Yok	Yok	Yok	Yok	Var
3. Soru	2	1	1	0	1	2	2	3

Örnek: 8 kişi üzerinde kan grupları üzerinde besin alerjisi araştırılmıştır.

	1	2	3	4	5	6	7	8
--	---	---	---	---	---	---	---	---

Kan grubu	O Rh+	ARh+	BRh+	BRh+	ORh-	A Rh-	ARh-	BRh-
Besin	Yok	Var	Var	Var	Yok	Yok	Var	Var

Bu tabloya göre;

- O gruplarda besin alerjisi yok.
- B gruplarda besin alerjisi bulunur.
- A grubunda besin alerjisi kesin olarak söylenemez.

sonuçları çıkar.

1.2. Tanım: Sayısal verilerin elde edildiği listeye istatistiksel seri denir, kısaca seri adı verilir. Yalnız seri kelimesini matematikteki seri ile karıştırmamak gerekir. Buradaki seri verilerin elde edildiği sayısal değerlerdir. Matematikte kullanılan diğer seri kavramı ise dizilerin terimlerinin toplamıdır.

Örnek: Bir ilçede bir ayda ölen insanların yaşları

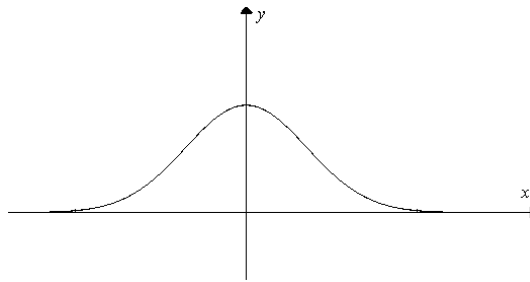
$$X = \{68, 73, 84, 2, 81, 48, 78, 79, 64, 68, 59\}$$

biçimindedir. Bu verilerin oluşturduğu listeye istatistiksel seri denir, kısaca seri adı verilir. //

Dikkat ederseniz bu veri listesine küme adı verilmemektedir. Çünkü küme-lerde bir eleman bir defa yazılır ama veri listesinde yani serilerde tekrar sayısı çok olabilir.

İstatistikte seriler $y = c.e^{-x^2}$, ($c \in \mathbb{R}$) fonksiyonunu türü şekiller oluşturur.

Örnek: $y = e^{-x^2}$ fonksiyonu aşağıdaki gibidir.



İstatistik bitimsel istatistik ve çözümsel istatistik olarak ikiye ayrılabilir.

1.3. Tanım: Araştırmaya ilişkin serilerinin toplanması, seriler yardımıyla gösterilmesi, sınıflandırılması ve grafikler yardımıyla gösterilmesi aşamalarını içeren sürece **betimsel istatistik** denir.

1.4. Tanım: Tablo ve grafikler yardımıyla gösterimi yapılan serilerinin amaca uygun farklı istatistik teknikler yardımıyla çözümlenmesi (analiz), sonuçların modeller yardımıyla ifade edilmesi ve ileriye dönük tahminlere (öngörü) bulunması süresine de **çözümsel istatistik** denir.

1.5. Tanım: Sayılabilir ve ölçülebilir özellikleri (değişkenleri) içeren, aralarında birçok benzerlikler olmakla beraber farklılıklar da bulunan nesnelere veya olaylara **istatistik birimi** denir. Sayılmayan veya ölçülmeyen nesnelere veya olaylar söz konusu olduğunda bunlar istatistik birimi oluşturmazlar. İstatistik birimi; canlı cansız bir olgu, bir olay veya bir kurum olabilir. Hastane, okul, fakülte ve insan gibi değerler istatistik birimini oluşturur. Ama koku, güzellik gibi ölçüsü belirlenmemiş değerlere istatistik birimi oluşturmaz.

1.6. Tanım: İstatistik birimlerine istenilen bir anda ortaya ulaşılabiliyorsa bu birime "**sürekli istatistik birimi**" denir. Ticari kuruluşlar, insanlar, otomobiller, fakülteler gibi serileri üzerine yapılan çalışmalar sürekli istatistik birimidir.

1.7. Tanım: İstatistik birimi belirli bir anda ortaya çıkan bu birime "**ani istatistik birimi**" denir. Yağmurlar, kavgalar, depremler, ölümler gibi serileri üzerine yapılan çalışmalar ani istatistik birimidir.

1.8. Tanım: Bir birimin alacağı değerler birimden birime farklılık gösterir. Bu farklılıklara değişken adı verilir. Örneğin bireyin mesleği, cinsiyeti, boyu, çocuklarının sayısı birer değişkenlerdir.

1.9. Tanım: Araştırma ilişkin tanımlanan istatistik birimlerinin tümünün oluşturduğu topluluğu **ana kitle (yığın, evren)** denir. Ana kitle istatistik biriminden oluştuğuna göre bunlardan farklı bir yapıya sahip olamaz. Örneğin, Ankara'daki bir askeri birlikteki askerler hakkında yapılacak bir araştırmada ana kitle, askeri birlikteki er ve erbaşların oluşturduğu topluluktur. Ana kitleyi oluşturan birimlerin toplam sayısı ana kitle çapı olarak bilinir ve N ile gösterilir. Ana kitle aynı zamanda, ana kitlede yer alan birimlerin ilgilenilen değişken veya değişkenler bakımından ölçülmesi ile elde edilen değerler topluluğudur. Ana kitlenin tamamının yerine, bu ana kitleden farklı tekniklerle oluşturulan daha az istatistik birimlerinden oluşan alt topluluklarla çalışır. Bu alt topluluğa da **örneklem** denir. Ana kitle şu şekilde çeşitleri vardır:

1. Soyut ve Somut Ana kitle
2. Gerçek ve Varsayılan Ana kitle
3. Hazır ve Hareketli Ana kitle

Şimdi bu tanımlar üzerinde duralım.

1.10. Tanım: Yapılacak çalışmalarda istatistik birimlerin tümüne ulaşılabiliyorsa bu birimlerden oluşan topluluğa **somut ana kitle** denir. İstatistik birimlerinin tümüne ulaşmanın olası olmadığı durumda oluşan topluluğa da **soyut ana kitle** denir.

1.11. Tanım: Gerçekte var olan istatistik birimlerinden oluşan kütleye **gerçek ana kitle** denir. Gerçekte var olmadığı veya ortaya çıkmadığı var olmaları ya da ortaya çıkmaları olası istatistik birimlerinden oluşan topluluğa da **varsayım kitle** denir. Kütüphanedeki kitap sayıları, üniversitedeki öğrenci sayıları, bir fakülte'deki öğretim üyesi sayısı, Tokat'taki trafik kazaları sayısı gibi veriler gerçek ana kitleye örnektir. Ama 10 sporcu arasında 6'şar sporcudan oluşan gruplar sayısı varsayım kitleye örnektir.

1.12. Tanım: Sürekli istatistik birimlerinden oluşan topluluğa **hazır ana kitle**, ani istatistik birimlerinden oluşan topluluğa **hareketli ana kitle** denir.

1.13. Tanım: Ana kitleden hesaplanan ve ana kitlenin özelliklerini tanımlayan matematiksel değerlere **parametre** denir. Örneğin, Ankara'daki askeri birlikteki er ve erbaşların boylarının aritmetik ortalaması bir parametredir.

1.14. Tanım: Ana kitleden seçilen ve ana kitleyi temsil edebilme özelliğine sahip birimlerin oluşturduğu topluluğa örnek adı verilir. Örneğin, Ankara'daki askeri birliklerden seçilen zırhlı birliklerin oluşturduğu topluluk örnek olacaktır. Örneğin oluşturan birimlerin toplam sayısı örnek çatı olarak bilinir ve n ile ifade edilir.

ORAN HIZI

1.15. Tanım: Bir serisindeki terimlerin yüzdelik oranlarına oran hızı denir.

Örnek: Bir toplulukta 35 erkek 52 bayan bulunmaktadır. Bu topluluktaki erkek oran hızı nedir?

$$\text{Çözüm: } \frac{35}{35+52} = 0,39$$

Bu topluluktaki erkek sayısı oran hızı % 39 dir.

Örnek: Bir hastane de 1 ay içerisinde 15 yaş üstü ölenlerin sayısı verilmiştir.

Yaş Aralığı	15-24	25-34	35-44	45-54	55-64	65-74	75-Üstü
Ölü Sayısı	1	3	5	8	23	52	47

Bu verilere göre 65-74 yaş aralığında ölüm oran hızı nedir?

Çözüm: Bu hastanede 1 ayda ölen sayısı 139 kişidir.

$$\frac{52}{139} = 0,374$$

65-74 yaş aralığında ölüm oran hızı % 37,4 dür.

VERİLERDE AÇIKLIK

1.16. Tanım: Bir serisinde en büyük değer ile en küçük değer arasındaki farka açıklık denir. Açıklığı R ile gösterilmek üzere,

$$R = \text{En büyük değer} - \text{En küçük değer}$$

dir.

Örnek: {20, 18, 25, 22, 19, 24} serisinin açıklığını bulunuz.

$$\text{Çözüm: } R = 25 - 18 = 7$$

VERİLERİN GRAFİKLE GÖSTERİMİ

İstatistiksel verilerin ve istatistiksel sonuçlarının gösterimi bazen grafiklerle yapılır. Grafikler;

i) Pasta (Daire) Grafiği (Pie Graphics)

1.17. Tanım: Gözlem sonucu elde edilen serileri, çevresi 360° olan daire üzeri de gösterilmesine pasta (daire) grafiği denir. Bir veri daire dilimine yerleştirilirken, serisinin tüm veriler toplamına oranı, verinin yerleştirileceği daire diliminin merkez açısının 360° ye oranına eşit olmalıdır.

Örnek: Bir lisede 72 erkek, 108 kız öğrenci bulunmaktadır. Bu okuldaki öğrencilerin cinsiyet durumunu pasta grafiği şeklinde gösteriniz.

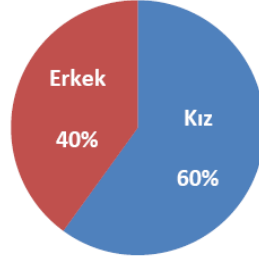
Çözüm: Toplam öğrenci sayısı 180 dir.

180 öğrenciden 72 si erkekse

360 öğrenciden x i matematikten geçmiştir.

$$x = \frac{72.360}{180} = 144^0$$

Kız öğrenci sayısının gösterimi $360 - 144 = 216^0$ dir.

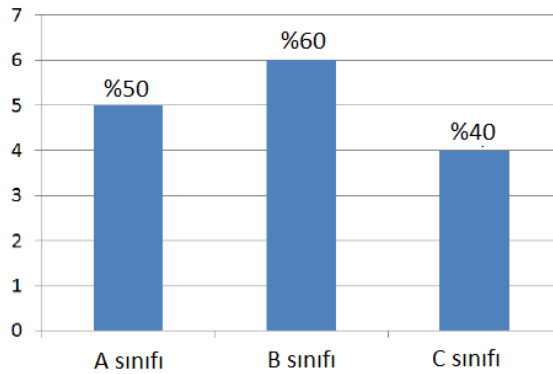


ii) Sütun Grafiği

1.18. Tanım: Gözlem sonucu elde edilen serileri, Kartezyen koordinat sisteminde sütun yükseklikleri ile gösterilmesine sütun grafiği denir.

Sütun grafiği dikey ve yatay olmak üzere iki biçimde kullanılabilir. Alışıldık olan gösterim biçimi dikey sütun grafiğidir. Bu grafikte, değişken ve değişkenin kategorileri yatay eksen ile veriler ise dikey eksen ile temsil edilir. Yatay sütun grafiğinde ise eksenlerin yerleri değişir.

Örnek: Bir okuldaki A, B ve C sınıflarında oluşan son sınıfa öğrencilerinden Çanakkale şehitliğini gezen öğrenciler sırasıyla % 50 , % 60 ve % 40 olduğu tespit edilmiştir. Bu sınıfların sütun grafiği aşağıdaki gibidir.



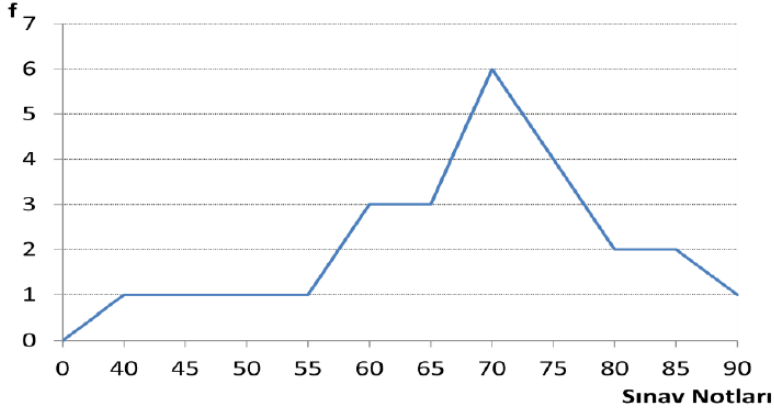
iii) Çizgi Grafiği

1.19. Tanım: Gözlem sonucu elde edilen serilerinin Kartezyen koordinatta parçalı fonksiyon olarak gösterimine çizgi grafiği denir.

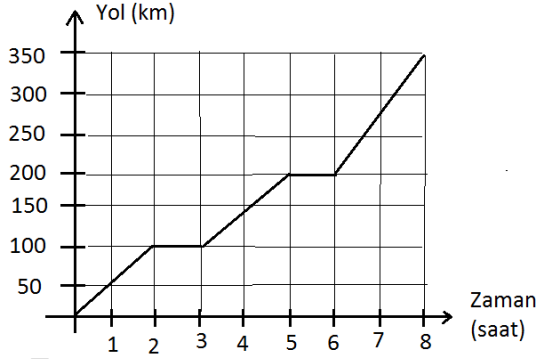
Örnek: Coğrafya dersinden alınan notların aşağıdaki tabloda verilmiştir.

Not	40	45	50	55	60	65	70	75	80	85	90
Frekans	1	1	1	1	3	3	6	4	2	2	1

Bu tablonun çizgi grafiği aşağıdaki gibidir.



Örnek: A şehrinden B şehrine giden bir aracın yol zaman grafiği aşağıdaki şekilde verilmiştir.



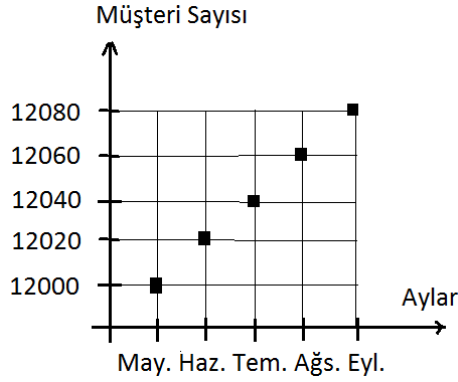
Bu şekilde 8 saatte 350 km yol almış, 2 ve 3 saatleri arasında 1 saat, 6 ve 7 saatleri arasında 1 saat olmak üzere mola vermiştir.

iv) Serpilme Grafiği

1.20. Tanım: x ve y iki değişken olmak üzere, (x,y) ikililerinin koordinat düzleminde belirttiği noktalardan oluşan grafiğe serpilme grafiği denir.

Örnek: Aşağıdaki tablo bir bankanın Mayıs-Eylül ayları arasında işlem yaptığı müşteri sayılarını göstermektedir?

Aylar	Mayıs	Haziran	Temmuz	Ağustos	Eylül
Müşteri Sayıları	12000	12020	12040	12060	12080



Şekilde görüldüğü gibi müşteri sayısı düzenli olarak artmaktadır.

1.1. Sonuç: i) Bir değişkenin bir bütün içerisindeki oranını belirlemek için daire grafiği,

ii) Bir değişkenin zaman içerisindeki değişimini incelemek için çizgi ya da sütun grafiği

iii) İki değişken arasındaki ilişkiyi göstermek için serpilme grafiğini kullanmak daha uygundur.

FREKANS, HİSTOGRAM ve DİYAGRAM

Frekans

1.21. Tanım: Herhangi bir konuda bir takım serisi elde edilebilir. Elde edilen verilere, istatistiksel işlemlerden geçirilmemişse buna ham veri adı verilir. Veriler reel olarak düzenlenir. Değişikliklere ilişkin bir ölçünün tekrar sayısına frekans (sıklık) adı verilir ve f harfi ile gösterilir.

Örnek: Bir sınıfın matematik dersinden almış olduğu notlar aşağıdaki şekildedir. Bu notları sıralayarak ve sınıflandırarak frekans dağılım biçiminde gösteriniz?

95	25	45	40	55	60	70	60	65	100
00	15	10	35	40	85	35	85	80	50
65	40	45	35	55					

Veri sayısı = 25 adettir.

Notların küçükten büyüğe doğru sıralanmış gösterimi;

00	10	15	25	35	35	35	40	40	40	45
45	50	55	55	60	60	65	65	70	80	85
85	95	100								

şeklinde dir. Frekans dağılımı iki şekilde olur. Bunlar:

1. Sınıflandırılmış (Sıralı) Frekans Dağılımı:

1.22. Tanım: Toplanan bütün verileri tek tek inceleyerek frekans oluşturmaya sınıflandırılmış (sıralı) frekans dağılımı denir.

Örnek: Yukarıdaki örneğin sınıflandırılmış (sıralı) frekans dağılımını gösterelim. Burada hangi nottan kaç kez tekrarlandığını almalıyız:

<u>Not</u>	<u>f_i</u>	<u>Not</u>	<u>f_i</u>	<u>Not</u>	<u>f_i</u>	<u>Not</u>	<u>f_i</u>
05	1	35	3	55	2	80	1
10	1	40	3	60	2	85	2
15	1	45	2	65	2	95	1
20	1	50	1	70	1	100	1

$\Sigma f_k = 25$ olup bu tabloya göre sıralı frekans sayısı $n = 16$ dir.

2. Gruplandırılmış Frekans Dağılımı:

1.23. Tanım: Toplanan bütün verileri belirli gruplarla inceleyerek frekans oluşturmaya gruplandırılmış frekans dağılımı denir.

Gruplandırılmış frekans dağılımında hangi sınıf aralığını tespit edilmesi gerekir. Bu sınıf aralığı şu şekilde tespit edilir.

1. Değer Aralığı; $R = \text{En Büyük Sayı} - \text{En Küçük Sayı}$
2. Sınıf Sayısı; $s \geq \sqrt{n}$ eşitsizliğini sağlayan yaklaşık doğal sayı
3. Sınıf Aralığı; $A \geq \frac{R}{s}$ eşitsizliğini sağlayan yaklaşık doğal sayı

Örnek: Yukarıdaki örneği önce sınıf aralığı formül ile bulup, gruplandırılmış frekans oluşturalım:

1. $R = 100 - 0 = 100$
2. $s \geq \sqrt{n} = \sqrt{25} = 5$
3. $A \geq \frac{R}{s} = \frac{100}{5} = 20$

Sınıf Aralığı	f_i
$0 \leq x < 20$	3
$20 \leq x < 40$	4
$40 \leq x < 60$	8
$60 \leq x < 80$	5
$80 \leq x \leq 100$	5

$\Sigma f_k = 25$ olup bu tabloya göre gruplandırılmış frekans sayısı $n = 25$ dir.

1.1. Aksiyom: Frekans dağılım tablosu üzerindeki her aralığa denk gelen sayıların yüzde oranı p ile gösterirsek gruplandırılmış frekans sayılarının yüzdelik durumu,

$$p_i = \frac{100}{n} f_i$$

şeklindedir.

Örnek: Yukarıda verilen gruplandırılmış frekans sayılarının yüzdelik durumu;

Sınıf Aralığı	f_i	$p(\%)$
$0 \leq x < 20$	3	$\frac{100}{n} f_i = \frac{100}{25} \cdot 3 = 12$
$20 \leq x < 40$	4	$\frac{100}{n} f_i = \frac{100}{25} \cdot 4 = 16$
$40 \leq x < 60$	8	$\frac{100}{n} f_i = \frac{100}{25} \cdot 8 = 32$
$60 \leq x < 80$	5	$\frac{100}{n} f_i = \frac{100}{25} \cdot 5 = 20$
$80 \leq x \leq 100$	5	$\frac{100}{n} f_i = \frac{100}{25} \cdot 5 = 20$

şeklindedir. Buna göre %12 öğrenci $0 \leq x < 20$ aralığında, % 16 öğrenci $20 \leq x < 40$, % 32 öğrenci $40 \leq x < 60$ ve % 20 öğrenci $60 \leq x < 80$ ve $80 \leq x \leq 100$ aralığında not almışlardır.

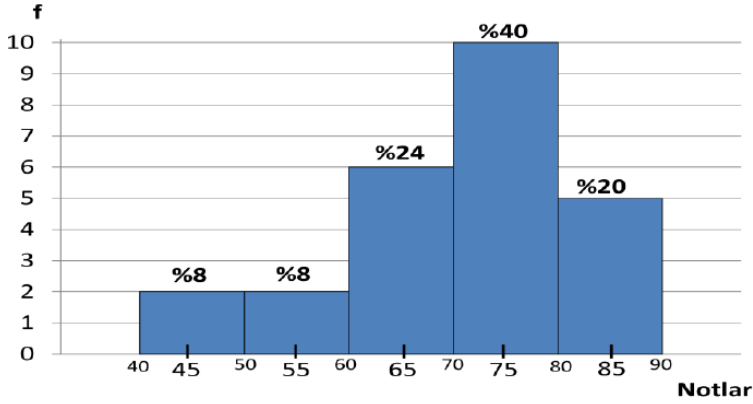
Histogram

1.24. Tanım: Gözlem sonucu elde edilen serisinin fazla olması durumunda tablo ve grafik oluşturmada karşılaşılan zorlukları ortadan kaldırmak için verileri gruplandırmak gerekir. Gruplandırılmış verilerle oluşturulan sütun grafiğe histogram denir.

Örnek: Bir sınıfta Türk Dili ve Edebiyatı dersinden alınan notlar ve frekansa sayıları aşağıdaki şekildedir.

Not Aralığı	Not Ortalaması	Frekansı (Alınan Not Sayısı)	Yüzdeler Oranı
40-49	45	2	% 8
50-59	55	2	% 8
60-69	65	6	% 24
70-79	75	10	% 40
80-90	85	5	% 20
Toplam		25	% 100

Bu sınıfın Türk Dili ve Edebiyatı dersinden alınan notlarının histogram grafiği aşağıdaki şekilde gibidir.



Diyagram

1.25. Tanım: Frekans dağılımında her sınıfının sınıf orta değerinin çizgi grafiğinde gösterimine **diyagram** denir. Burada sınıf orta değeri \bar{x}_j olmak üzere;

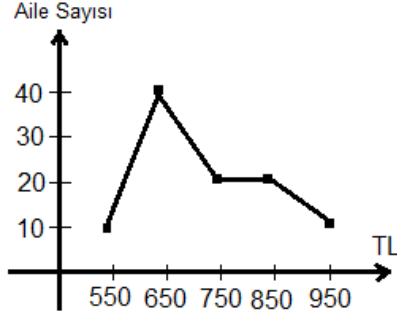
$$\bar{x}_j = \frac{j \text{ sınıfının alt değeri} + j \text{ sınıfının üst değeri}}{2}$$

dir.

Örnek: 100 ailenin aylık gıda harcamalarının ilişkin frekans dağılımı aşağıdaki gibi olsun.

Gıda Harcaması (TL)	Aile Sayısı (f_i)	Sınıf Orta Değeri (\bar{x}_j)
$500 \leq x < 600$	10	550
$600 \leq x < 700$	40	650

$700 \leq x < 800$	20	750
$800 \leq x < 900$	20	850
$900 \leq x \leq 1000$	10	950
N	100	



Ailelerin gıda harcamaları diyagramı şekilde gibidir.

Öğr. Gör. Şaban YILMAZ