İSTATİSTİK
DERS NOTLARI

<u>HAZIRLAYANLAR</u>

YDR.DOÇ.DR.ANDIM OBEN BALCE - YRD.DOÇ.DR.SERDAR DEMİR

PAMUKKALE ÜNİVERSİTESİ

İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ

EKONOMETRİ BÖLÜMÜ

2007, DENİZLİ

BÖLÜM 1 GİRİŞ

Günümüzde artık neredeyse her gün duyduğumuz "İSTATİSTİK" sözcüğü acaba gerçekte neyi ifade etmektedir? Özellikle medyada her gün bir kısım istatistiksel bilgiler sunulmakta, doğru ya da yanlış yorumlar yapılmaktadır. Ekonomik analizlerde, spor programlarında, seçim süreçlerinde, sağlık konularında ve daha birçok alanda sıkça kullanılan istatistik neden bu kadar önemlidir? Aklımıza gelen bu sorulara, bu bölümde kısaca cevap verilmeye çalışılacak, istatistiğin temel kavramları ele alınarak bir giriş yapılacaktır.

İSTATİSTİK (STATISTICS) NEDİR?

Yaşantımızın her anında bir karar verme durumu söz konusudur. Bu kararları verirken sahip olduğumuz bilgileri düşünce sürecinden geçirerek sonuca varmaya çalışırız. "Peki sahip olduğumuz bilgiler her zaman yeterli midir?, Bu bilgileri nasıl elde ediniriz?" v.b. sorulara vereceğimiz cevaplar bizi istatistiğe yönlendirecektir. İhtimallerin bu kadar çok olduğu yaşamda elbetteki alınacak kararların, içinde bulunulan koşulların en iyi biçimde değerlendirilerek alınması gerekmektedir. Bu değerlendirme sürecinde en güçlü yardımcı araçlardan birisi istatistiktir. İstatistik, matematiğe dayalı olan kuramını uygulamaya geçirerek her türlü alanda büyük katkılar sağlamaktadır. Günümüzde ekonomi, işletme, sağlık, spor, mühendislik, genetik, astronomi, sanat v.d. hemen hemen tüm alanlarda kullanılan istatistiksel yöntemler, hızla artan bir öneme sahiptir. Bu nedenle istatistiğin doğru biçimde anlatılması gerekmektedir. Burada temel düzeyde istatistik konuları ele alınarak, öğrencilere günlük yaşamlarında ya da iş yaşamlarında yararlı olacak bilgileri vermek amaçlanmaktadır .

Bir tanım olarak istatistik; belirsizlik altında bir konuda karar verebilmek amacıyla, ilgilenilen konuya ilişkin verilerin toplanması, düzenlenmesi, özetlenmesi, çözümlenmesi ve sonuçlarının yorumlanmasına yönelik olarak kullanılan yöntemler topluluğu olarak ifade edilebilir. Kısacası veri bilimidir ya da verilerin dilidir.

ÖNEMLİ KAVRAMLAR, TERİMLER, TANIMLAR

Araştırma (Research)

İlgilenilen konuya ilişkin sorunların saptanması, çözüm yollarının planlanması, uygulamaya konulması ve sonuçlarının değerlendirilmesine yönelik yapılan çalışmalardır. Araştırmanın konusu ve amacı belirgin olmalıdır. Zaman, personel ve maliyet dikkate alınarak, araştırmanın sınırı iyi belirlenmelidir. Araştırmada görev alan araştırıcılar ya da uzmanlar yeterli bilgi düzeyine sahip olmalıdır. Araştırmanın sonucunda elde edilen bilgiler doğru biçinde değerlendirilmelidir.

Ölçme ve Ölçü (Ölçek)

Ölçme (Measurement): Araştırma konusu ile ilgili sayısal değerleri elde etme işlemine " ölçme" denir. Ölçme başlı başına bir çalışma alanıdır. Birçok yöntem ve teknik kullanılmaktadır.

Ölçek(Scale): Sayısal değerleri elde etmek için kullanılan araç ya da gereçlere "Ölçek" ya da "Ölçek ya da Ölçü" denir. Ölçek türleri aşağıda verilmektedir:

Sınıflamaya dayalı (Nominal) Ölçek: Cinsiyet (E, K); Sektör (Otomotiv, İmalat, Tarım, Maden,...; Medeni Durum; v.b; Ülkeler; ...)

Sıralamaya dayalı (Ordinal) Ölçek: Ünvanlar, Rütbeler, Sınıflar,...

Aralıklı (Interval) Ölçek: Başlangıç ve bitiş noktası vardır. Geçme notları, Sıcaklık ölçüleri, Zeka Ölçekleri,...

Orana dayalı (Ratio) Ölçek: Sabit bir başlangıç noktası vardır. Hacim ve ağırlık ölçüleri, Uzaklık ölçüleri, sermaye, ...

Kitle (Yığın, Anakütle, Population)

Araştırma kapsamına giren, aynı özellikleri taşıyan birimlerin ya da bireylerin oluşturduğu topluluğa KİTLE denir. Kitlenin büyüklüğü araştırmanın özelliğine göre değişir. Nüfus sayımı için kitle Türkiye'dir. Denizli'deki üniversite öğrencilerinin giderleri için kitle Pamukkale Üniversitesi öğrencileridir. Kitle büyüklüğüne bağlı olarak her zaman tüm birimler (bireyler) hakkında bilgi sahibi olmak mümkün değildir. Bundan dolayı geniş kitlelerde araştırmalar; zaman, maliyet, personel, ulaşım, vb. nedenlerden dolayı tüm birimler yerine daha az sayıdaki birimler seçilerek yürütülür.

Örneklem (Sample) ve Örnekleme (Sampling)

Bir kitleden, belirli yöntemler kullanılarak seçilen aynı özellikleri taşıyan bir kısım bireyin oluşturduğu topluluğa ÖRNEKLEM denir.
Örneklem seçmek için kullanılan yöntemler topluluğu ise ÖRNEKLEME olarak adlandırılır. Örnekleme yöntemleri en genel şekliyle, Olasılığa Bağlı ve Olasılığa Bağlı Olmayan olarak 2 grubta toplanabilir.

Olasılığa Bağlı Bazı Örnekleme Yöntemleri: Basit Rasgele Örnekleme, Tabakalı Örnekleme, Sistematik Örnekleme, Küme Örneklemesi; Sıralı Küme Örneklemesi.

Olasılığa Bağlı Olmayan Bazı Örnekleme Yöntemleri: Kota Örneklemesi, Kartopu Örneklemesi, Uzman Örneklemesi.

Tam Sayım

Bir araştırma kapsamında, kitledeki tüm birimlerine ulaşılarak istenen bilginin elde edilmesi işlemidir. Bunun yapılabilmesi için incelenecek kitlenin büyüklüğünün, belirlenen maliyet ve zaman gibi kısıtlara uygun olması gerekir. Bazı durumlarda (nüfus sayımları gibi) kitle büyük olsa bile tam sayım yapılması zorunlu olmaktadır. Gelişen teknoloji ile birlikte bu tür tam sayımlar daha kolay yapılabilir hale gelmiştir.

Gözlem (ya da denek, Observation)

Kitle ya da örneklemde yer alan her birime gözlem ya da denek denir. Gözlem (ya da denek) sayısı aşağıdaki biçimde simgeleştirilmektedir.

Kitledeki Gözlem Sayısı : **N**
Örneklemdeki Gözlem Sayısı: **n**

Parametre (Parameter) ve İstatistik

Kitle özelliklerinin sayısal değerlerine PARAMETRE denir. Araştırma kitle yerine örneklem üzerinde uygulanıyorsa, parametre değerleri tahmin edilir. Bu durumda, örneklemden elde edilen sayısal değerlere İSTATİSTİK denir. Örnek olarak, sıkça kullanılan bazı parametreler ve istatistikler aşağıda verilmektedir:

Parametre	***İstatistik***
Kitle Ortalaması : μ	Örneklem Ortalaması : $\bar{x}$
Kitle Varyansı : σ^2	Örneklem Varyansı : S^2

Değişken (Variable)

Nicel (kantitatif) ya da nitel (kalitatif) anlamda bir özellik ya da karakterde belirgin olarak görülen farklılık, DEĞİŞKEN ile gösterilebilir. Bir değişken, denekten deneğe değişebilir. Örneğin; GSMH, İhracat-İthalat Değerleri ya da miktarları, Cari Fiyat, Dolar Alış-Satış Kurları,vb...

Değişkenlere karşılık gelen denek ya da gözlem değerlerine de VERİ denir. Veriler, **tek değişkenli**, çift değişkenli ya da **çok değişkenli** olarak da kullanılabilir.

Değişkenler; **Nicel** ve **Nitel** değişkenler olarak sınıflandırılabilir.

Örnek:	***Nitel Değişkenler***	***Nicel Değişkenler***
	Cinsiyet	İhracat Miktarı
	Medeni durum	Öğrenci Sayısı
	Şirket Türü	TÜFE Endeksi
	Ülkeler	Ücretler

Değişkenler; **Sürekli** ve **Kesikli** değişkenler olarak sınıflandırılabilir.

Sürekli Değişken: Değişkenler ölçülerek ya da sıralanarak elde edilir. İki ölçüm arası sonsuz sayıda noktaya bölünebilir. Aralık biçiminde ifade edilebilirler. Örneğin; boy uzunluğu, kilo, fiyat, gelir,....

Kesikli Değişken: Ölçümler 0, 1, 2 gibi kesin değerler alır. Ara değerler söz konusu değildir. Nitel değişkenler, genellikle kesikli değişkenlerdir. Örneğin; cinsiyet, yazı-tura, ülke kodları, şirket türü, vb...

BÖLÜM 2 VERİLERİN ÖZETLENMESİ

Bu bölümde, veri kavramı ve verilerin özetlenmesinde kullanılan temel yöntemler ele alınacak; çeşitli uygulamalar ile konuların daha iyi anlaşılması sağlanmaya çalışılacaktır.

Veri (Data, Datum)

Belirli amaçlar için toplanan bilgilere veri denir. Veri toplamak için farklı yöntemler kullanılabilir:

- Mevcut kaynaklardan (eski kayıtlar, arşivler, raporlar, yıllıklar, vb...) yararlanarak,
- Gözlem yaparak,
- Anket yaparak,
- Deney yaparak,
- Simülasyon yoluyla bilgisayarda yapay veri üreterek.

Temel olarak 2 tür veri vardır:

NİCEL (Quantity) VERİ: Sayısal bir ölçekle ölçülerek elde edilmiş verilerdir. Boy uzunluğu, fiyat,..

NİTEL (Qualify) VERİ: Kategoriler biçiminde sınıflandırılabilen verilerdir. Cinsiyet, göz rengi,...

Dağılım (Distribution)

Kitlede ya da örneklemde yer alan her değişkene ilişkin veriler, araştırma konusuna ve araştırılan topluluğa özgü bir dağılım gösterirler. Dağılımda yer alan verilerin ortalaması hesaplanabilir. Veriler, ortalamanın iki yanında ve farklı uzaklıklarda yer alırlar. Verilerin, ortalamanın iki yanında ne şekilde yer aldığına ilişkin görünüme verilerin DAĞILIMI denir. İstatistikte kullanılan dağılım kavramı genel olarak 3 başlık altında ele alınmaktadır:

i) Sıklık Dağılımları
ii) Olasılık Dağılımları
iii) Örneklem Dağılımları

Bunlar ilerleyen kısımlarda ya da bölümlerde ayrıntılı olarak ele alınacaktır.

SIKLIK DAĞILIMLARI

Çeşitli yollarla toplanan veriler, özellikleri hakkında bilgi edinmek amacıyla, düzenlemeler yapılarak özet halinde sunulurlar. Uygulamalarda veri kümeleri çok sayıda gözlem içerebilmektedir. Bu nedenle verilerin özetlenmesi, ilgilenilen olay ya da problem açısından ilk yapılacak iş ve son derece önemli bir işlemdir. Verilerin düzenlendiği çizelgelere ***sıklık çizelgeleri***, verilerin gösterdiği dağılıma ***sıklık dağılımı*** denir. Verilerin yapısına (nitel, nicel, vb...) göre sıklık çizelgeleri düzenlenir.

NİCEL VERİLERDE SIKLIK ÇİZELGELERİ

Bir araştırma sonunda elde edilen sürekli nicel veriler, düzenlenmemiş ham ya da sınıflandırılmamış verilerdir. Aşağıda örnek olarak sunulan veriler düzenlenmemiş ham verilerdir. Konunun daha iyi anlaşılması için bu örnek üzerinden uygulamalar yapılacaktır.

ÖRNEK: Bir finans analisti, bilgisayar donanım ve yazılım şirketlerinin Araştırma-Geliştirme(AR-GE) faaliyetlerine ayırdıkları kaynak miktarıyla ilgilenmektedir. Bu analist yüksek teknolojiye sahip 50 firmayı örneklem olarak belirlemiş ve bir önceki yıl gelirlerinden AR-GE'ye ayırdıkları kaynak miktarlarını (1000 YTL) elde etmiştir. Analistin amacı bu veri kümesini özetleyerek bir takım bilgilere ulaşmaktır. Veriler Tablo 1'de verilmiştir.

Tablo1. Firmaların AR-GE faaliyetlerine ayırdıkları kaynak miktarları

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
1	13.5	**11**	8.0	**21**	8.2	**31**	9.6	**41**	7.1
2	8.4	**12**	7.9	**22**	8.0	**32**	7.2	**42**	13.2
3	10.5	**13**	6.8	**23**	7.7	**33**	8.8	**43**	7.7
4	9.0	**14**	9.5	**24**	7.4	**34**	11.3	**44**	5.9
5	9.2	**15**	8.1	**25**	6.5	**35**	8.5	**45**	5.2
6	9.7	**16**	13.5	**26**	9.5	**36**	9.4	**46**	5.6
7	6.6	**17**	9.9	**27**	8.2	**37**	10.5	**47**	11.7
8	10.6	**18**	6.9	**28**	6.9	**38**	6.9	**48**	6.0
9	10.1	**19**	7.5	**29**	7.2	**39**	6.5	**49**	7.8
10	7.1	**20**	11.1	**30**	8.2	**40**	7.5	**50**	6.5

Bu verilerin kaynak miktarına göre küçükten büyüğe doğru sıralanmış şekli, Tablo2'de verilmektedir.

Tablo2. Sıralanmış Kaynak Miktarı Verileri

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
45	5.2	**28**	6.9	**43**	7.7	**35**	8.5	**9**	10.1
46	5.6	**38**	6.9	**49**	7.8	**33**	8.8	**3**	10.5
44	5.9	**10**	7.1	**12**	7.9	**4**	9.0	**37**	10.5
48	6.0	**41**	7.1	**11**	8.0	**5**	9.2	**8**	10.6
25	6.5	**29**	7.2	**22**	8.0	**36**	9.4	**20**	11.1
39	6.5	**32**	7.2	**15**	8.1	**14**	9.5	**34**	11.3
50	6.5	**24**	7.4	**21**	8.2	**26**	9.5	**47**	11.7
7	6.6	**19**	7.5	**27**	8.2	**31**	9.6	**42**	13.2
13	6.8	**40**	7.5	**30**	8.2	**6**	9.7	**1**	13.5
18	6.9	**23**	7.7	**2**	8.4	**17**	9.9	**16**	13.5

Sıklık Çizelgesinin Elde Edilişi

Sıklık çizelgesinin elde edilişinde kullanılan tanımlar adımsal olarak verilerek yukarıdaki örnekle bu tanımlar pekiştirilmeye çalışılacaktır.

Dağılım Sınırları: Bir dağılımda (veri kümesinde) yer alan en küçük ve en büyük denek değerleridir.

En büyük değer (Maksimum): 13.5 (dağılımın üst sınırı)

En küçük değer (Minimum) : 5.2 (dağılımın alt sınırı)

Dağılım Genişliği (DG): Dağılım sınırları arasındaki farktır.

DG=En büyük değer - En küçük değer =13.5 - 5.2 = 8.3

Sınıf: Eşit ya da birbirine yakın değerli deneklerin oluşturduğu her bir gruba SINIF denir. Sınıf sayısı, **k** ile gösterilir. Sınıf sayısının genellikle 7 ile 20 arasında olması istenir. Araştırmacı tarafından belirtilen sınıf sayısı, çok sayıda veri olduğunda aşağıda verilen Sturges'in formülü ile de bulunabilir.

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

Sınıfın Alt Sınırı: Bir sınıfta yer alan en küçük değerdir.

Sınıfın Üst Sınırı: Bir sınıfta yer alan en büyük değerdir.

Sınıf Aralığı: Ard Arda gelen iki sınıfın alt sınırları ya da üst sınırları arasındaki farktır. Sınıf aralığı, c ile gösterilir. Örneğimizde sınıf sayısı 8 olarak alınsın. Bu durumda Sınıf Aralığı:

$$c = \frac{DG + a}{SINIFSAYISI}$$

a: veri kümesindeki verilerin ondalık kısmındaki hane sayısı ile ilgilidir. Örneğimizde tam kısımdan sonra 1 hane olduğu için a=0.1 alınır.

$$c = \frac{8.3 + 0.1}{8} = \frac{8.4}{8} = 1.05$$

Sınıf sayısı ve aralığı belirlendikten sonra, **ilk sınıfın alt sınırı saptanır**. Bu değer genellikle dağılımın en küçük değeridir. Sınıf aralığı (c=1.1) ard arda eklenerek **diğer sınıfların alt sınırları** bulunur. **İlk sınıfın üst sınırı** ise ikinci sınıfın alt sınırının **son hanesinden** 1 çıkarılarak bulunur. Diğer sınıfların üst sınırları da ardı ardına sınıf aralığı eklenerek bulunur.

Sıklık: Bir sınıfta yer alan denek sayısı o sınıfın sıklığıdır. f ile gösterilir. Sıklıklar toplamı denek sayısına eşittir.

$$\sum_{i=1}^{n} f_i = n$$

Örneğimizde ise k=8 tane sınıf olduğundan $\sum_{i=1}^{8} f_i = 50$'dir.

Sınıflar oluşturulduktan sonra sınıfların sıklıkları bulunur. Bunun için önce her sınıfın sıklığı işaretleme (çeteleme) ile bulunur. İşaretlerin sayısı **sınıf sıklıklarını** verir.

Seçilen **sınıf sayısının uygun olup olmadığını** tespit etmek için son sınıfta en büyük denek değerinin yer alıp almadığın bakılmalıdır. Eğer son sınıfta en büyük denek değeri yer almıyorsa sınıf sayısı daha büyük olmalıdır. En büyük denek değeri son sınıftan bir önceki sınıfta da yer alabilir. Bu durumda seçilen sınıf sayısı veriler için büyüktür. Bu iki durumda sınıf sayısı artırılır ya da azaltılır.

Sınıf (Orta) Değeri (m): Bir sınıfın alt ve üst sınırlarının ortalaması o sınıfın sınıf değeri ya da sınıf orta değeridir. Sınıf değeri bir sınıfı tek bir değerle temsil eder ve m ile gösterilir.

$$m_i = \frac{AS_i + ÜS_i}{2}, \quad i = 1,2,...,k$$

AS_i: i.sınıfın alt sınırı

$ÜS_i$: i.sınıfın üst sınırı

Göreli Sıklık(Sıklık Yüzdesi): Her sınıfa düşen denek sayısının toplam denek sayısına göre yüzdesidir. Göreli Sıklıklar p_i ile gösterilir. Toplamları 1 olmalıdır.

$$p_i = \frac{f_i}{n} \quad , i = 1,2,...,k \qquad \qquad \sum_{i=1}^{n} p_i = 1$$

Çizelge1. AR-GE faaliyetleri için ayrılan kaynak miktarı verileri için sıklık çizelgesi

Sınıf	Alt Sınır (AS)	Üst Sınır (ÜS)	Sınıf Orta Değeri (m_i)	Çeteleme	Sıklık (f_i)	Göreli Sıklık (p_i=f_i/n)
1	5.20	6.24	(5.15+6.24)/2=5.72	////	4	0.08
2	6.25	7.29	6.77	///// ///// //	12	0.24
3	7.30	8.34	7.82	///// ///// ///	13	0.26
4	8.35	9.39	8.87	/////	5	0.10
5	9.40	10.44	9.92	///// //	7	0.14
6	10.45	11.49	10.97	/////	5	0.10
7	11.50	12.54	12.02	/	1	0.02
8	12.55	13.59	13.07	////	3	0.06
				Toplam:	**50**	**1.00**

Çizelge 1 için örnek yorumlar:

İkinci sınıfta AR-GE'ye 6.25 ile 7.29 bin YTL arasında yatırım yapan firmalar yer alır ve bu sınıftaki firmalar ortalama 6.77 bin YTL AR-GE'ye yatırım yapar.

AR-GE'ye yatırım yapan 50 firmanın 12 tanesi ya da %24'ü AR-GE'ye ortalama 6.77 bin YTL yatırım yapmışlardır.

AR-GE'ye yatırım yapan firmaların %24'ü AR-GE'ye ***tahminen*** ortalama 6.77 bin YTL yatırım yapmaktadırlar. (kitle için yorum)

Sınıf Ara Değerleri (SA): Sınıflar arasındaki değerlerdir. Birinci sınıfın üst sınırı ile ikinci sınıfın alt sınırının ortalaması, birinci sınıf ile ikinci sınıf arasındaki sınıf ara değerini verir. İlk sınıfın ara değeri birinci sınıftan önce bir sınıf varmış diye kabul edilerek, hesaplanın birinci ve ikinci sınıflar arasındaki değerinden sınıf aralığı (c) çıkarılarak bulunur. Son sınıf ara değeri ise, sanki son sınıftan sonra bir sınıf daha varmış gibi kabul edilerek, hesaplanan son sınıf ara değerine sınıf aralığı (c) eklenerek bulunur. Sınıf ara değerlerinin sayısı (k+1)'dir.

Birikimli Sıklık: Sınıf sıklıklarının üst üste eklenmesi ile oluşan sıklıklar birikimli sıklıklardır.

Den Daha Az Birikimli Sıklık: *Sınıf ara değerinden* daha az değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir.

Den Daha Çok Birikimli Sıklık: *Sınıf ara değerinden* daha çok değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir.

Bir sınıf ara değerine karşı gelen den daha az ve den daha çok birikimli sıklıklar toplamı denek sayısına eşittir. Birikimli sıklıklar denek sayısına oranlanırsa **den daha çok birikimli**

sıklık yüzdeleri ve **den daha çok birikimli sıklık yüzdeleri** elde edilir. *Birikimli sıklık yüzdeleri sınıf ara değerinden daha az ya da daha çok büyük değerli denek değerlerinin yüzdesini verir.*

Alt sınır, üst sınır, sınıf değeri, sıklık ve göreli sıklık bilgilerinin oluşturduğu çizelgeye Sıklık Çizelgesi (Çizelge 1) ve Alt sınır, üst sınır; sınıf ara değeri, birikimli den daha az ve den daha çok sıklıklar ve birikimli den daha az ve den daha çok sıklık yüzdelerinden oluşan çizelgeye Birikimli Sıklık Çizelgesi denir(Çizelge 2).

Çizelge2. AR-GE faaliyetleri için ayrılan kaynak mikatarı verileri için Birikimli Sıklık Çizelgesi

Sınıf	AS	ÜS	Sınıf Ara Değeri	f_i	Birikimli Sıklıklar		Birikimli Sıklık Yüzdeleri	
					Den Daha Az	Den Daha Çok	Den Daha Az	Den Daha Çok
			5.145		0	50	0.00	1.00
1	5.20	6.24		4				
			6.245		4	46	0.08	0.92
2	6.25	7.29		12				
			7.295		16	34	0.32	0.68
3	7.30	8.34		13				
			8.345		29	21	0.58	0.42
4	8.35	9.39		5				
			9.345		34	16	0.68	0.32
5	9.40	10.44		7				
			10.445		41	9	0.82	0.18
6	10.45	11.49		5				
			11.495		46	4	0.92	0.08
7	11.50	12.54		1				
			12.545		47	3	0.94	0.06
8	12.55	13.59		3				
			13.595		50	0	1.00	0.00

Çizelge 2 için örnek yorumlar:

50 firmadan sadece 4 tanesi ya da %8'i AR-GE'ye 6.245 bin YTL'**den daha az** yatırım yapmıştır.

AR-GE'ye yatırım yapan firmalardan tahmini olarak %8'i AR-GE'ye 6.245 bin YTL'**den daha az** yatırım yapmaktadır. (kitle için yorum)

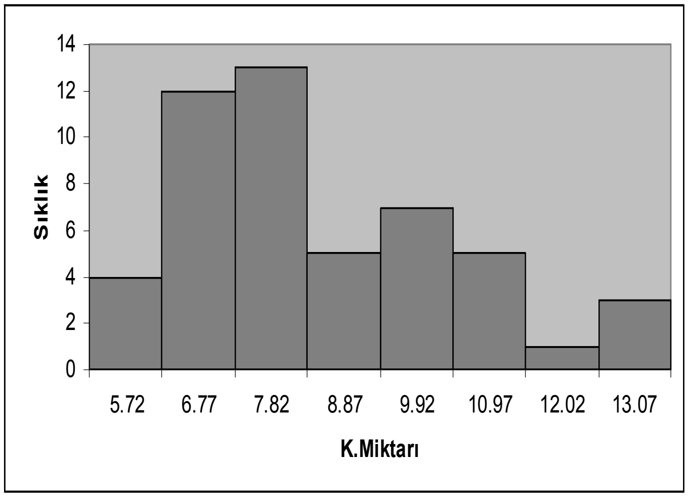
50 firmadan sadece 46 tanesi ya da %92'si AR-GE'ye 6.245 bin YTL'**den daha çok** yatırım yapmıştır.

AR-GE'ye yatırım yapan firmalardan tahmini olarak %92'i AR-GE'ye 6.245 bin YTL'**den daha çok** yatırım yapmaktadır. (kitle için yorum)

Dikkat: Rakamların geldiği yerlere dikkat ederek benzer yorumlar diğer sınıflar içinde yapılabilir.

Veriler ondalıklı değil ise sınıf ara değerinin anlamı yoktur. Çizelge2'deki sınıf ara değeri kolonu olmaz ve birikimli sıklık ve birikimli sıklık yüzdeleri alt sınır ve üst sınır değerlerine göre yorumlanır.

HİSTOGRAM :



Çizelge1'de verilen göreli sıklık ve sınıf değerleri kullanılarak histogram çizilmiş ve yukarıda verilmiştir. Tablodaki bilgiler görsel olarak bu histogramdan da yorumlanabilir. Histogramdan veri dağılımının sağa çarpık olduğu gözlenmektedir. Büyük olasılık ile verilerin ortalaması ortancadan büyüktür. Eğer ortalama ve ortanca hesaplanırsa bu kolaylıkla görülebilir.

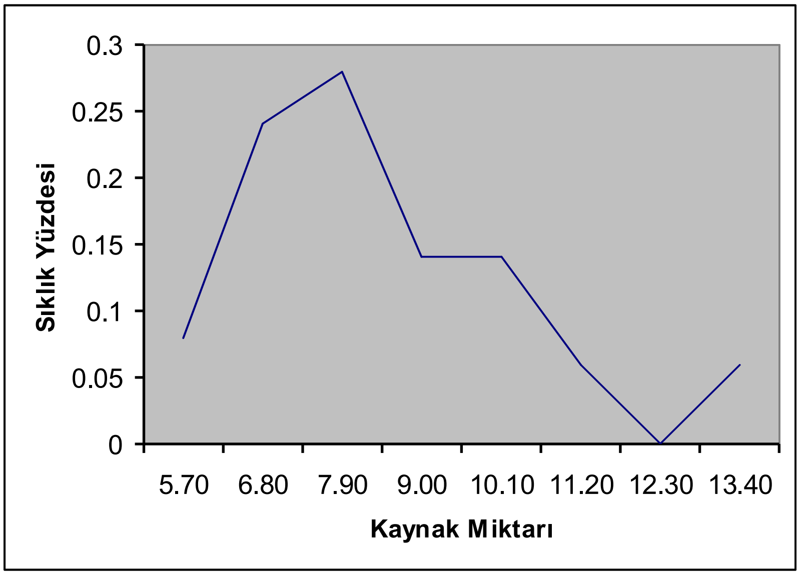
DAL YAPRAK ÇİZİMİ:

Veri kümesinin görsel olarak bir diğer gösterimi Dal-yaprak çizimidir.

Dal	Yaprak
5	2 6 9
6	0 5 5 5 6 8 9 9 9
7	1 1 2 2 4 5 5 7 7 8 9
8	0 0 1 2 2 2 4 5 8
9	0 2 4 5 5 6 7 9
10	1 5 5 6
11	1 3 7
12	
13	2 5 5

Dal kısmında ondalıktan önceki verinin tam kısmı, yaprak kısmına ise ondalıktan sonraki kısım yazılarak dal-yaprak çizimi gerçekleştirilir. Benzer yorumlar bu çizimden de yapılır.Veri kümesinin sağa çarpıklığı buradan da gözlenebilir. 50 firmanın çoğunluğunun %9.9'dan daha az AR-GE'ye yatırım yaptığı söylenebilir. Bu durumdaki firma sayısı ise 40 tanedir. Grafikten yorumlar subjektif'tir. Diğer benzer ya da farklı yorumlamak mümkündür.

DAĞILIM POLİGONU:



Enterpolasyon yöntemi ile ara değer bulma

Yukarıdaki örnekte size %9'dan daha çok yatırım firma sayısı ne olduğu sorulduğunda tablodan direk bu soruya cevap vermeniz mümkün değil ama enterpolasyon yöntemi ile cevap vermeniz mümkündür.

Sınıf Ara Değeri	Den Daha çok Sıklığı
8.345	21
9.00	X
9.395	16

$$\frac{(9-8.345)}{(9.395-8.345)} = \frac{X-21}{16-21} \Rightarrow \frac{0.655}{1.05} = \frac{X-21}{-5} \Rightarrow X \cong 17.88 \cong 18$$

Enterpolasyon yöntemi ile %9'dan daha çok yatırım firma sayısının yaklaşık 18 olduğu bulunur. Enterpolasyon yöntemi ara değer bulmada kullanılan bir yöntemdir.

NİTEL VERİLERDE SIKLIK ÇİZELGESİ

SINIFLANABİLEN VERİLERDE SIKLIK ÇİZELGESİ

Sınıflanan verilerde, sınıflar bağımsız olarak elde edildiği için her sınıfa düşen denek sayıları sıklık çizelgesini oluşturur. Sınıflar bağımsız olduğu için sıklık çizelgesinde sadece sınıf ve sıklık, göreli sıklık kolonları yer alır.

ÖRNEK: 2003 turizm sezonunda ülkemize gelen 800 turistin ülkeye gelişte yararlandıkları taşıt türlerine göre dağılımının sıklık çizelgesi aşağıdaki çizelgede verilmiştir.

Çizelge 3. 800 turistin yararlandıkları taşıt türlerine göre sıklık çizelgesi

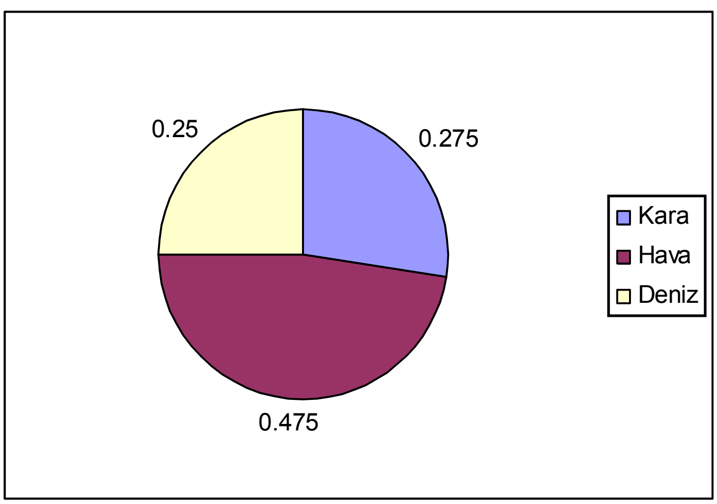
Taşıt Türü	Sıklık (f_i)	Göreli Sıklık ($p_i=f_i/n$)
Kara Taşıtı	220	0.275
Hava Taşıtı	380	0.475
Deniz Taşıtı	200	0.250
Toplam	800	1.000

Yorumlar:

2003 turizm sezonunda ülkemize gelen 800 turistin %47.5'i (380 tanesi) uçak ile seyahat etmişlerdir.

2003 turizm sezonunda ülkemize gelen turistlerin yaklaşık %47.5'inin uçak ile seyahat ettikleri tahmin edilmektedir.

Daire Dilimleri Grafiği



SIRALANABİLEN VERİLERDE SIKLIK ÇİZELGESİ

Veriler belli bir sıralama ölçütüne göre sıralanabilen sınıflara ayrılır ve her sınıfa düşen deneklerin sayısı saptanırsa sıralanabilen verilerde sıklık çizelgesi elde edilir.

ÖRNEK: Ekonomi dersinin genel sınav sonuçlarının sıklık çizelgesi aşağıdaki çizelgede verilmiştir.

Çizelge 4. 77 öğrencinin Ekonomi sınav sonuçlarının sıklık çizelgesi

Notlar	Sıklık (f_i)	Göreli Sıklık ($p_i=f_i/n$)	Ve Daha Az Birikimli Sıklığı	Ve Daha Çok Birikimli Sıklığı
F	29	0.377	29 (%37.7)	77 (%100)
C	31	0.402	60 (%77.9)	48 (%62.3)
B_2	7	0.091	67 (%87.0)	17 (%22.0)
B_1	6	0.071	73 (%94.8)	10 (%12.9)
A_2	3	0.039	76 (%98.7)	4 (%5.2)
A_1	1	0.013	77 (%100)	1 (%1.3)
	77	**1.000**		

Yukarıdaki sıklık çizelgesinden yararlanarak bazı yorumlar yapılabilir: Ekonomi dersini alan öğrencilerin %3.9'u dersi A_2 notu ile başarmıştır. Ekonomi dersini alan öğrencilerin %62.3'ü C **ve daha yüksek** not almıştır

EŞİT OLMAYAN ARALIKLI VE AÇIK UÇLU SIKLIK ÇİZELGESİ

Eşit olmayan aralıklı sıklık çizelgeleri, uçlara doğru sıklık yoğun olduğunda ya da dağılım aşırı derecede çarpıklaştığında, uçların birinde ayrıntıların yok olduğu diğerinde gereksiz olduğu durumlarda düzenlemelidir. Örneğin, gelir dağılımı eşit olmayan aralıklı çizelgesi oluşturulur. Düşük gelirlerde sıklık yayılması olmasına karşın yüksek gelirlerde sıklıklar azalır.

Eşit olmayan aralıklı sıklık çizelgelerinde, sınıflandırmada yorum ve grafik çiziminde kolaylık sağlanması için aralıklar en küçük aralığın katları olarak alınmalı ve değişik aralık sayısı da az olmalıdır.

Dağılım sınırları belli olmayan verilerin sıklık çizelgeleri açık uçlu düzenlenir. Açık uçlu sıklık çizelgeleri, sınıf değerlerine bağlı olan hesaplamalarda güçlük yaratabilir. Böyle durumlarda kesin olmayan sınıf değerleri tahmin edilir. Açık uçlu sınıf sıklıkları küçük olduğu için, tahmin edilen değerlerden ötürü hatalı sonuç bulma olasılığı küçük olabilir.

ÖRNEK: Bir ildeki 30 ilkokulun derslik sayılarının dağılımı incelensin.

Derslik Sayısı:	5 6 7 8 9 10 11 12 13 14 15
İlkokul Sayısı:	6 8 0 1 0 9 0 3 0 2 1

Verilerin, derslik sayısı 7, 9 11 ve 13 olan ilkokul sayısı sıfır olduğu için eşit olmayan aralıklı sıklık çizelgesi düzenlenebilir:

Derslik Sayısı	İlkokul Sayısı
5	6
6	8
7-9	1
10-12	12
13-15	3

ÖRNEK: 108 kişinin yetenek test puanlarına göre dağılımı, açık uçlu sıklık çizelgesine örnektir:

Puan	Sıklık(f_i)
100'den az	7
100-119	11
120-139	24
140-159	36
160-179	19
180-199	8
200 ve çok	3

MERKEZİ KONUM (EĞİLİM) ÖLÇÜLERİ

Konum ölçüleri, verilerin dağılımdaki yerlerini, birbirlerine olan uzaklıklarını kısacası konumlarını belirlemek için kullanılan ölçülerdir. Bu ölçüler Tablo1'de verilen verilen veriler üzerinde uygulanarak aşağıda tek tek ele alınmaktadır.

Aritmetik Ortalama ($\bar{x}$)

En sık biçimde kullanılan merkezi konum ölçüsüdür.

Sınıflandırılmamış verilerde aritmetik ortalama,

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

formülü ile bulunur. Elimizdeki verilerin aritmetik ortalaması;

$$\bar{x} = 5.2 + 5.6 + ... + 13.5 = 424.6 / 50 = 8.49$$

olarak bulunur.

Sınıflandırılmış verilerde ise aşağıdaki formüller kullanılmaktadır:

i) Formül1

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i m_i}{n} = (5.72x4+6.80x12+...+13.07x3)/50 = 425.65/50 = 8.513$$

ii) Formül2

$$\bar{x} = A + c\frac{\sum_{i=1}^{k} f_i b_i}{n}$$

A=Herhangi bir sınıf değeri (genellikle orta kısımlardaki sınıflardan seçilir)

$b_i=(S_i-A)/c$

Tepe Değeri (Mod) ($\hat{x}$)

Sınıflandırılmamış verilerde tepe değeri en sık tekrar eden değerdir. Ham verilerimizde bu değer 6.50 ve 8.20 değerleridir.

Sınıflandırılmış verilerde ise tepe değeri aşağıdaki gibi hesaplanır.

$$\hat{x} = A_s + c\frac{d_1}{d_1 + d_2}$$

A_s=En büyük sıklığın bulunduğu sınıfın alt sınırı

d_1=En büyük sıklık - bir önceki sıklık

d_2=En büyük sıklık-bir sonraki sıklık

O halde,

A_s=7.30

d_1=13-12=1

d_2=13-5=8

$$\hat{x} = 7.30 + 1.05\frac{1}{1+8} \cong 7.42$$

<u>Ortanca (Medyan)</u>($\bar{x}'$ **)**

Sınıflandırılmamış verilerde ortanca; n çift ise n/2'nci değer, n tek ise (n+1)/2'nci değerdir. Veri sayısı çift olduğundan (50)/2=25'inci değer olan 8.00 ortanca değerimizdir.

Sınıflandırılmış verilerde ise ortanca aşağıdaki gibi hesaplanır.

$$\bar{x}' = L + c\frac{\frac{n(\text{yada } n+1)}{2} - \sum_{m=1}^{i-1} f_m}{f_i}$$

$$j = \begin{cases} \frac{n}{2} & , n \text{ çift ise} \\ \frac{n+1}{2} & , n \text{ tek ise} \end{cases}$$

L= DDA kolonunda j'nin bulunduğu sınıfın alt SA_i değeri

n=50 olduğundan j=59/2=25 bulunur.

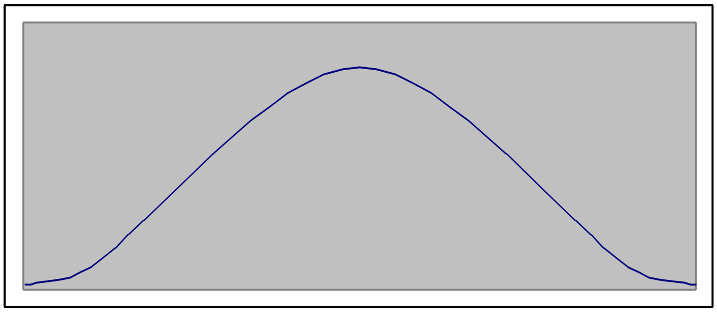
L=7.295

$$\sum_{m=1}^{i-1} f_m = \sum_{m=1}^{4-1} f_m = 16$$

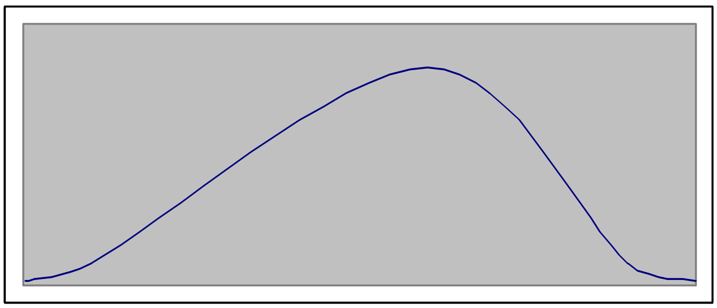
$$\bar{x}' = L + c\frac{n/2 - \sum_{m=1}^{i-1} f_m}{f_i} = 7.295 + 1.05\frac{50/2 - 16}{13} \cong 8.02$$ olarak bulunur.

Ortalama, ortanca ve tepe değeri arasındaki bağıntı:

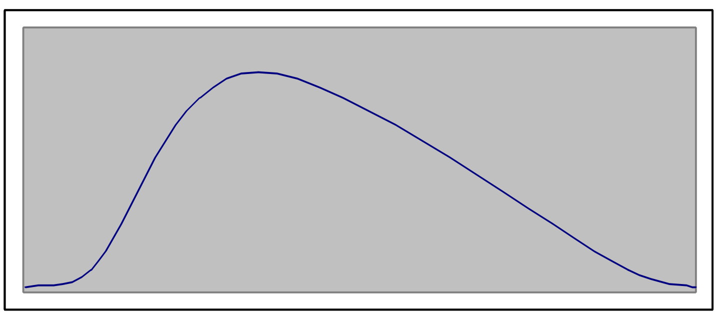
$\bar{x} = \bar{x}' = \hat{x}$ ise sıklık dağılımı simetrikdir.



$\bar{x} < \bar{x}' < \hat{x}$ ise sıklık dağılımı negatif yöne eğilimli ya da sola doğru çarpıktır.



$\bar{x} > \bar{x}' > \hat{x}$ ise sıklık dağılımı pozitif yöne eğilimli ya da sağa doğru çarpıktır.



Verilerimiz için $\hat{x} < \bar{x}' < \bar{x}$ olduğundan dağılımın sağa çarpık olduğu söylenebilir. Hali hazırda verilerin histogram grafiğinden de bu görünmekteydi.

YÜZDELİKLER, ÇEYREK DEĞERLER (DÖRDEBÖLENLER)

Yüzdelik (Percentile): Kendinden önce ve sonraki belirli oranlarda değerler olan noktanın değerini ifade etmektedir. Örneğin; 40. yüzdelik kendisinden önce deneklerin %40'ını kendisinden sonra deneklerin % 60'ının olduğu değerdir.

Çeyrek Değer (Quartile): 25. ,50. ve 75. yüzdelik değerleri, 1, 2 ve 3. çeyrek dğerlerdir.

Q_1= 1.çeyrek değer (25. yüzdelik) (Alt dördebölen)
Q_2= 2. çeyrek değer (50. yüzdelik) aynı zamanda ortancadır.
Q_3= 3. çeyrek değer (75. yüzdelik) (Üst dördebölen)

Sınıflandırılmamış verilerde aşağıdaki eşitlikler ile 1. ve 3. çeyrek değerler bulunur:

$$Q_1 = \begin{cases} X_j & , j = \dfrac{n+1}{4} \text{ , n tek ise} \\ \dfrac{X_j + X_{j+1}}{2} & , j = \dfrac{n}{4} \text{ , n çift ise} \end{cases} \qquad Q_3 = \begin{cases} X_j & , j = \dfrac{3(n+1)}{4} \text{ , n tek ise} \\ \dfrac{X_j + X_{j+1}}{2} & , j = \dfrac{3n}{4} \text{ , n çift ise} \end{cases}$$

n ya da n+1 dördün katı değilse, çeyrek değerler araya katma yolu bulunurlar.

Verilerimiz için n=50 olduğundan dördün katı değildir. j=50/4=12.50=$12^{1/2}$ olduğundan 12. ile 13. veri arasında kalmaktadır (% 50 lilik bir fazlayla). Bu durumda;

Q_1=6.9+0.50x(7.1-6.9)=7.00

3.çeyrek değeri için j=3x50/4=37.5=$37\frac{1}{2}$ olduğundan 37 ile 38. veri arasında kalmaktadır. Bu durumda;

Q_3=9.6+0.50x(9.7-9.6)=9.65

Dördebölenler Aralığı: Aşırı uç değerden çok az etkilenen bir konum ölçüsüdür. Ancak kullanımda tercih edilmeyen bir ölçüdür.

DBA=Q_3- Q_1

Sınıflandırılmış verilerde ise aşağıdaki formül ile hesaplanmaktadır:

$$P_a = L + c\frac{\%an - \sum_{m=1}^{i-1} f_m}{f_i}$$

P_a=a. yüzdelik

L=%an değerini içine alan DDA sıklıklarının minimum olanına karşılık gelen SA değeri

Σf_m=%an değerini içine alan DDA sıklığı

f_i=yüzdeliğin bulunduğu sınıfın sıklığı

Verilerimiz için aşağıdaki yüzdelikleri bulalım:

a=10 için

%an=10x50/100=5.00, L=6.245, f_2=12, Σf_m=4

$$P_{10} = 6.245 + 1.1\frac{5-4}{12} \cong 6.34$$

YORUM-1: 50 firmanın **%10'u** AR-GE için **en fazla 6.34 bin YTL** yatırım yapmaktadır.

YORUM-2: 50 firmanın **%90'ı** AR-GE için **en az 6.34 bin YTL** yatırım yapmaktadır.

YORUM-3: Firmaların **%10'u** AR-GE için **tahminen en fazla 6.34 bin YTL** yatırım yapmaktadır.

YORUM-4: Firmaların **%90'ı** AR-GE için **tahminen en az 6.34 bin YTL** yatırım yapmaktadır.

a=25 için

%an=25x50/100=12.50 L=6.245 f_2=12 Σf_m=4

$$P_{25} = 6.245 + 1.1\frac{12.5-4}{12} \cong 7.02$$

YORUM-1: 50 firmanın **%25'i** AR-GE için **en fazla 7.02 bin YTL** yatırım yapmaktadır.

YORUM-2: 50 firmanın **%75'i** AR-GE için **en az 7.02 bin YTL** yatırım yapmaktadır.

YORUM-3: Firmaların **%25'i** AR-GE için **tahminen en fazla 7.02 bin YTL** yatırım yapmaktadır.

YORUM-4: Firmaların **%75'i** AR-GE için **tahminen en az 7.02 bin YTL** yatırım yapmaktadır.

a=50 için

%an=50x50/100=25 L=7.295 f_3=13 Σf_m=16

$$P_{50} = 7.295 + 1.05\frac{25-16}{13} \cong 8.02$$

YORUM-1: 50 firmanın **%50'si** AR-GE için **en fazla 8.02 bin YTL** yatırım yapmaktadır.

YORUM-2: 50 firmanın **%50'si** AR-GE için **en az 8.02 bin YTL** yatırım yapmaktadır.

YORUM-3: Firmaların **%50'si** AR-GE için **tahminen en fazla 8.02 bin YTL** yatırım yapmaktadır.

YORUM-4: Firmaların **%50'si** AR-GE için **tahminen en az 8.02 bin YTL** yatırım yapmaktadır.

a=75 için

%an=75x50/100=37.50 L=9.395 f_5=7 Σf_m=34

$$P_{75} = 9.395 + 1.05\frac{37.5 - 34}{7} \cong 9.92$$

YORUM-1: 50 firmanın **%75'i** AR-GE için **en fazla 9.92** bin YTL yatırım yapmaktadır.

YORUM-2: 50 firmanın **%25'i** AR-GE için **en az%9.92** bin YTL yatırım yapmaktadır.

YORUM-3: Firmaların **%75'i** AR-GE için **tahminen en fazla 9.92** bin YTL yatırım yapmaktadır.

YORUM-4: Firmaların **%25'i** AR-GE için **tahminen en az 9.92** bin YTL yatırım yapmaktadır.

AĞIRLIKLI, GEOMETRİK VE HARMONİK ORTALAMA

AĞIRLIKLI ORTALAMA

Sınıflandırılmamış bazı veri kümelerinde verilerin önem dereceleri farklı olabilir. Bu farkların etkisi de ağırlık biçiminde hesaplamaya katılarak Ağırlıklı Ortalama elde edilir. Ağırlıklı ortalama aşağıdaki eşitlik ile hesaplanmaktadır. X_i değerini alan f_i tane denek olmak üzere,

$$AO = \sum_{i=1}^{k} f_i X_i \Big/ \sum_{i=1}^{k} f_i$$

Sıklık tablosundan elde edilen ortalama da bir tür ağırlıklı ortalamadır. Benzer bir örnek:

ÖRNEK: Bir öğrencinin bir dönem boyunca aldığı derslere ilişkin ders kredisi ve not değerleri aşağıda verilmektedir.

Ders Adı	Kredisi	Notu
Matematik I	4	B1
Okuma Becerileri	2	A2
Kimya I	4	B2
Fizik I	4	C2
Türk Dili I	2	B1

Buna göre bu öğrencinin dönem ağırlıklı ortalaması;

AO=(4x3+2x3.5+4x2.5+4x1.5+2x3)/18=2.277 olarak bulunur.

ÖRNEK: Bir iş yerinde çalışan 20 işçinin günlük ücreti 2500 YTL, 33 işçinin 3000 YTl, 40 işçinin 3200 YTL ve 10 işçinin de 4000 YTL'dir. Bu iş yerinde ortalama ücreti hesaplayınız.

AO=(20x2500+33x3000+40x3200+10x4000)/100=3080 YTL olarak bulunur.

GEOMETRİK ORTALAMA

Geometrik ortalama ölçümler arasında değişme oranı çok olduğunda hesaplanan bir ortalamadır. Genellikle nüfus büyüme hızı, gelişme hızı ya da endeks gibi hesaplamalarda kullanılır. Geometrik ortalama uç değerlerden aritmetik ortalama kadar etkilenmez. Herhangi bir denek değeri 0 ya da negatif olduğunda geometrik ortalama hesaplanamaz.

Sınıflandırılmamış verilerde geometrik ortalama

$$GO = \sqrt[n]{X_1 X_2 ... X_n} = (X_1 X_2 ... X_n)^{1/n}$$

formülü ile hesaplanmaktadır.

Sınıflandırılmış verilerde ise $GO = \sqrt[n]{S_1^{f_1} S_2^{f_2} ... S_n^{f_n}} = (S_1^{f_1} S_2^{f_2} ... S_n^{f_n})^{1/n}$ ile elde edilmektedir.

ÖRNEK: Bir bölgenin nüfusu 1980 yılında 4254670 kişi, 1982 yılında ise 4575470 kişi ise 1982 yılındaki nüfusu tahmin ediniz.

$$GO = \sqrt{4254670.4575470} = 4412153.4$$

Geometrik Ortalama ile Bileşik Faiz Formülü Arasındaki İlişki

Geometrik ortalamanın hesaplandığı eşitlikten yaralanılarak bileşik fazi formülüne ulaşılabilir. Bunu aşağıdaki örnekle gösterelim.

ÖRNEK: Bir malın toptan satış fiyatı 1.yıldan 2.yıla %40, 2.yıldan 3.yıla ise % 50 oranında artmıştır. İncelenen 2 yıl içindeki ortalama artış ne kadardır?

1.yıl satış fiyatı: 1 birim ise
2.yıl satış fiyatı: 1.4 birim,
3.yıl satış fiyatı: 2.1 birim olur.

$$GO = \sqrt{1.40x1.50} \cong 1.45$$

Gözlenen 2 yıl içinde malın toplam satış fiyatındaki ortalama artış yaklaşık % 45'dir.

Geometrik ortalama formülünden yararlanarak bileşik faiz formülüne ulaşılabilir.

$$\sqrt{1.40x1.50} \cong 1.45$$

$$1.40x1.50 \cong (1+0.45)^2$$

$$2.10 \cong (1+0.45)^2$$

Bileşik faiz formülü,

$$P_t = P_0\ (1+r)^t$$

P_0: Temel değer

P_t: t zaman sonraki değer

r: Artış oranı

t: zaman

Buradan örneğimiz için P_0= 1.00, P_t= 2.10, r = 0.45, t = 2 olur.

Harmonik Ortalama

Genellikle ortalama hız, ortalama fiyat v.b. hesaplamalarda kullanılmaktadır: Aşağıdaki formül ile hesaplanır.

$$H.O = \frac{1}{\frac{1}{n}(\frac{1}{X_1} + \ldots + \frac{1}{X_n})}$$

ÖRNEK: Bir şehirler arası otobüs gittiği mesafesinin ilk üçte birinde 300km/s, ikinci üçte birinde 450 km/s ve son üçte birinde 360 km/s hız yapmıştır. Buna göre aracın ortalama hızı ne olmuştur.

$$H.O = \frac{1}{\frac{1}{3}(\frac{1}{300} + \frac{1}{450} + \frac{1}{360})} = 360\,\text{km/s}$$

DEĞİŞİM (YAYILIM) ÖLÇÜLERİ

Konum ölçüleri veri kümesinin ya da dağılımının merkezi hakkında bilgi varmektedir. Yalnızca bu ölçülere bakılarak verinin dağılımı hakkında tam bir bilgi sahibi olmak mümkün değildir. Veriler ne derecede dağılmakta ya da yayılım göstermektedirler?; ortalamadan uzaklıkları ne kadardır? Gibi sorulara cevap vermek için değişim ölçüleri hesaplanmalıdır.

ÖRNEK: NEWBOLD (s.16). 2 firmanın 7 ustabaşısının yıllık kazançları aşağıda veilmiştir.

A FİRMASI:	34500	30700	32900	36000	34100	33800	32500
B FİRMASI:	34900	27500	31600	39700	35200	33800	31700

İki firmaya ilişkin ortalama ve ortanca değerleri hesaplandığında eşit bulunmaktadır.

Ortalama= 33500 ve Ortanca=33800.

Konum ölçülerine baktığımızda 2 firma arasında bir farklılık görünmüyor. Gerçekte verilerin yayılımına dikkat ettiğimizde, A firmasında çalışanların birbirlerine daha yakın kazanç elde ettiklerini, buna karşın B firmasındakilerin ise kazançların birbirlerinden daha uzak olduğunu görmekteyiz. B firmasına çalışanların kazançları daha fazla farklılık göstermektedir. Az sayıdaki bu verilere bakarak bu sonuca varabiliyoruz. Ancak veri sayısının fazla olduğu durumda, bakarak bir sonuça ulaşmamız imkansızlaşacaktır.

DEĞİŞİM GENİŞLİĞİ (ARALIK): Bir veri kümesindeki en büyük değer ile en küçük değer arasındaki farktır. Değişim genişliği **R (RANGE)** harfi ile gösterilir.

$$R= X_{maks.}- X_{min.}$$

ÇEYREK SAPMA: Ortalam yerine ortanca kullanıldığında ya da aşırı uç değerler bulunduğunda değişim genişliği yerine ÇEYREK SAPMA (Q) kullanılır.

$$Q=\frac{Q_3-Q_1}{2}=\frac{9.62-7.02}{2}=1.3$$

Dağılımdaki tüm değerler kullanılmadığı için bilgi kaybı olabilir.

ORTALAMA MUTLAK SAPMA: Verilerin ortalamadan farklarının (sapmalarının) mutlak değerlerinin ortalamasıdır.

$$MS=\frac{\sum_{i=1}^{n}|X_i-\bar{x}|}{n}$$

formülü ile hesaplanır.

VARYANS ve STANDART SAPMA

Bir veri dağılımındaki değişimin önemli bir ölçüsü varyanstır. Varyansın karekökü alınarak standart sapma elde edilir.

Sınıflandırılmamış verilerde varyans ve standart sapma formülleri ve verilerimiz için sonuçları:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{192.2168}{49} \cong 3.84$$ **Kitle için**: $$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

S=1.96

S= Standart Sapma $\qquad$ S^2=Varyans

Örneklem için neden (n-1)'e bölünüyor ?

20 koltuk bulunan bir salona öğrenciler sırasıyla girsinler. İlk öğrencinin 20 koltuktan birini seçme serbestliği var. İkinci öğrencinin kalan 19 koltuktan birini seçme serbestliği var. Bu şekilde devam edilirse, 19. öğrencinin kalan 2 koltuktan birini seçme serbestliği var. Ancak 20. öğrencinin koltuk seçme serbestliği kalmamıştır. Demek ki 19 öğrencinin seçme serbestliği varken 1 öğrencinin yoktur.

Sınıflandırılmış verilerde ise aşağıdaki formüller yardımıyla hesaplanır.

$$S^2 = \left(\frac{\sum_{i=1}^{k} f_i m_i^2 - (\sum_{i=1}^{k} f_i m_i)^2 / n}{n-1} \right)$$

$$S^2 = \frac{3816.739 - (425.65)^2 / 50}{50-1} \cong 3.94$$ S=1.98

ÖRNEK: NEWBOLD ustabaşı örneği.

A Firması için $\sigma^2 = 2414286$, σ=1554

B Firması için $\sigma^2 = 12368571$, σ=3517

DEĞİŞİM, ÇARPIKLIK VE BASIKLIK KATSAYISI

Değişim Katsayısı

Standart sapma ortalamanın yüzdesi olarak tanımlandığında "Değişim Katsayısı" olarak adlandırılır. Özellikle farklı ölçü birimleri kullanan iki dağılımın değişimlerinin karşılaştırılmasında kullanılır. Aşağıdaki eşitlik ile hesaplanmaktadır.

$$V = \frac{S}{\bar{X}} 100 \qquad V = \frac{1.98}{8.49} 100 = 23$$

Çarpıklık ve Basıklık

Dağılımın ortalamaya göre biçimine ilişkin bazı bilgileri çarpıklık ve basıklık katsayıları ile öğrenebiliriz.

Çarpıklık katsayısı aşağıda verilen eşitlik ile elde edilebilir.

$$ÇK = \frac{\sum_{i=1}^{n}(X_i - \overline{x})^3 / n}{S^3}$$

ÇK=0 ise dağılım simetriktir.

ÇK<0 ise dağılım sola doğru çarpık ya da - yöne eğilimlidir.

ÇK>0 ise dağılım sağa doğru çarpık ya da + yöne eğilimlidir.

Basıklık Katsayısı ise aşağıdaki eşitlik ile hesaplanabilir:

$$B = \frac{\sum_{i=1}^{n}(X_i - \overline{x})^4 / n}{S^4} - 3$$

B=0 ise dağılımın yüksekliği standart normal dağılıma uygundur.

B<0 ise dağılım standart normalden basıktır.

B>0 ise dağılım standart dağılımdan sivridir.

Verilerimizden elde edilen çarpıklık ve basıklık katsayısı değerleri:

$$ÇK = \frac{\sum_{i=1}^{n}(X_i - \overline{x})^3 / n}{S^3} \cong 0.80$$

$$B = \frac{\sum_{i=1}^{n}(X_i - \overline{x})^4 / n}{S^4} - 3 \cong 3.13 - 3 \cong 0.13$$

Verilerimiz sağa çarpık ve standart dağılımdan daha sivri bir dağılıma sahiptir.

UYGULAMA-1

Aşağıda 45 ailenin yıllık gelirleri (1000 YTL) verilmektedir. Bu verilerden yararlanılarak sıklık çizelgesi, birikimli sıklık çizelgesi, konum ölçüleri, değişim ölçüleri elde edilmektedir.

12.4	13.7	27.5	19.4	12.5	28.3	36.3	29.7	19.7
19.6	32.6	9.9	21.6	41.2	36.4	19.8	23.4	9.6
41.7	30.4	11.4	7.5	38.7	20	38.2	11.3	25.6
17.3	26.8	22.1	20.7	24.3	9.9	23.2	35.8	14.2
21.9	23.8	16.4	24.7	21.5	22.2	29.8	29.3	21.4

Verileri küçükten büyüğe doğru sıralayalım.

7.5	*9.6*	*9.9*	*9.9*	*11.3*	*11.4*	*12.4*	*12.5*	*13.7*
14.2	*16.4*	*17.3*	*19.4*	*19.6*	*19.7*	*19.8*	*20*	*20.7*
21.4	*21.5*	*21.6*	*21.9*	*22.1*	*22.2*	*23.2*	*23.4*	*23.8*
24.3	*24.7*	*25.6*	*26.8*	*27.5*	*28.3*	*29.3*	*29.7*	*29.8*
30.4	*32.6*	*35.8*	*36.3*	*36.4*	*38.2*	*38.7*	*41.2*	*41.7*

En Küçük (Min.)= 7.5

En Büyük (Max.)= 41.7

Dağılım Genişliği (Range)= Max. - Min. = 41.7 - 7.5 = 34.2

k (sınıf sayısı) = 7 alınırsa

c (sınıf aralığı) = (34.2+0.1) / 7= 34.3/ 7 = 4.9 olarak bulunur.

Sıklık çizelgesini düzenleyelim:

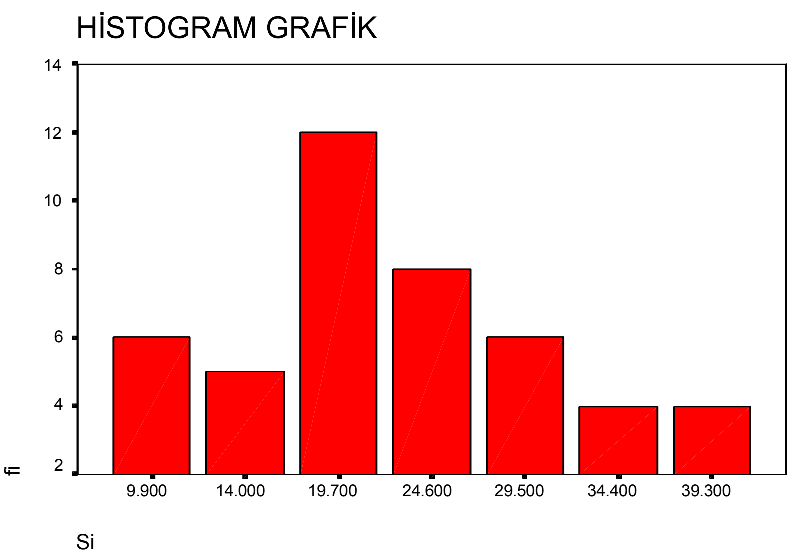
Alt Sınır(A.S)	*Üst Sınır(Ü.S.)*	*Sınıf Değeri (m_i)*	*Sıklık (f_i)*	*Göreli Sıklık (p_i)*	*bi* .
7.5	12.3	9.9	6	0.133	-3
12.4	17.2	14.8	5	0.111	-2
17.3	22.1	19.7	12	0.267	-1
22.2	27	24.6	8	0.178	0
27.1	31.9	29.5	6	0.133	1
32	36.8	34.4	4	0.089	2
36.9	41.7	39.3	4	0.089	3

m_i (sınıf değeri)= (AS_i + $ÜS_i$) / 2

p_i (göreli sıklık)= f_i / n

Birikimli sıklık çizelgesini düzenleyelim:

A.S.	***Ü.S.***	f_i	m_i	SA_i	***D. D.A.***f_i	***D.D.Ç.***f_i	***D. D.A.***p_i	***D. D.Ç.***p_i
				7.45	0	45	0.000	1.000
7.5	12.3	6	9.9					
				12.35	6	39	0.133	0.867
12.4	17.2	5	14.8					
				17.25	11	34	0.244	0.756
17.3	22.1	12	19.7					
				22.15	23	22	0.489	0.511
22.2	27	8	24.6					
				27.05	31	14	0.689	0.311
27.1	31.9	5	29.5					
				31.95	37	8	0.822	0.178
32.0	36.8	4	34.4					
				36.85	41	4	0.911	0.089
36.9	41.7	4	39.3					
				41.75	45	0	1	0



Yıllık geliri 25 Bin YTL'den fazla olan aile sayısını bulmak isteyelim. Bunu interpolasyon yardımıyla yapabiliriz.

SA_i	D.D.Ç.f_i
22.15	22
25	x
27.05	14

$$\frac{25-22.15}{27.05-22.15}=\frac{x-22}{14-22}$$ eşitliği ile x aşağdaki gibi bulunur.

$$\frac{2.85}{4.9}=\frac{x-22}{-8}$$ $x \cong 17$ bulunur.

Şimdi yıllık geliri 15 Bin YTL'den az olan aile sayısını bulalım.

SA_i	D.D.A.f_i
12.35	6
15	x
17.25	11

$\frac{17.25-12.35}{15-12.35}=\frac{11-6}{x-6}$ eşitliği ile x aşağdaki gibi bulunur.

$$\frac{4.9}{2.65}=\frac{5}{x-6}$$

$x \cong 9$ bulunur.

ORTALAMA, TEPE DEĞERİ (MOD) VE ORTANCA (MEDYAN)

Ortalama ($\bar{x}$)

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i s_i}{n} = (9.9x6+14.8x5+...+19.7x12)/45 = 1038.4/45 = 23.075$$

Tepe Değeri ($\hat{x}$)

$$\hat{x}=A_s+c\frac{d_1}{d_1+d_2}$$

A_s=17.3	En büyük sıklığın bulunduğu sınıfın alt sınırı
d_1=12-5=7	En büyük sıklık - bir önceki sıklık
d_2=12-8=4	En büyük sıklık-bir sonraki sıklık

$$\hat{x}=17.3+4.9\frac{7}{7+4}=17.3+3.12=20.42$$

Ortanca ($\bar{x}'$)

$$\bar{x}'=L+c\frac{\frac{n(\,yada\;n+1)}{2}-\sum_{m=1}^{i-1} f_m}{f_i} \qquad j=\begin{cases}\frac{n}{2} & ,\text{n çift ise}\\ \frac{n+1}{2} & ,\text{n tek ise}\end{cases}$$

L= DDA kolonunda j'nin bulunduğu sınıfın alt SA_i değeri

n=45 olduğundan j=46/2=23 bulunur.

L=17.25

$$\sum_{m=1}^{i-1} f_m=\sum_{m=1}^{4-1} f_m=11$$

$$\bar{x}'=L+c\frac{n/2-\sum_{m=1}^{i-1} f_m}{f_i}=17.25+4.9\frac{46/2-11}{12}=21.95 \text{ olarak bulunur.}$$

Ortalama, ortanca ve tepe değeri arasındaki bağıntı

23.175 > 21.95 > 20.42 olduğundan dağılım sağa doğru çarpıktır.

YÜZDELİKLER, ÇEYREK DEĞERLER

$$P_a = L + c\frac{\%an - \sum_{m=1}^{i-1} f_m}{f_i}$$

P_a=a. yüzdelik

L=%an değerini içine alan DDA sıklıklarının minimum olanına karşılık gelen SA değeri

Σf_m=%an değerini içine alan DDA sıklığı

f_i=yüzdeliğin bulunduğu sınıfın sıklığı

<u>a=25 için</u>

%an=25x45/100=11.25
L=17.25
f_3=12
Σf_m=11

$$P_{25} = 17.25 + 4.9\frac{11.25 - 11}{12} = 17.35$$

<u>a=50 için</u>

%an=50x45/100=22.5
L=17.25
f_3=12
Σf_m=11

$$P_{50} = 17.25 + 4.9\frac{22.5 - 11}{12} = 21.95$$

<u>a=75 için</u>

%an=75x45/100=33.75
L=27.05
f_5=6
Σf_m=31

$$P_{75} = 27.05 + 4.9\frac{33.75 - 31}{6} = 29.29$$

VARYANS, STANDART SAPMA

$$S^2 = \left(\frac{\sum_{i=1}^{k} f_i s_i^2 - (\sum_{i=1}^{k} f_i s_i)^2 / n}{n - 1}\right)$$

$$S^2 = \left(\frac{(27314.52 - 1038.4^2 / 45}{45 - 1}\right) = 76.201$$

S=8.73

ALIŞTIRMALAR

1) Bir benzin istasyonu rasgele seçtiği 30 müşterisinin aldığı benzin miktarlarını litre olarak aşağıdaki gibi kaydetmiştir. Sıklık ve birikimli sıklık çizelgelerini 7 sınıf için düzenleyiniz. Ortalama, ortanca, tepe değeri ve varyansını çizelgelerden yararlanarak hesaplayınız. (Cula, s.76)

2	2	3	4	6	7	7	8	8	9
10	10	10	11	11	11	11	12	12	12
13	13	14	16	18	20	25	25	28	28

2) İktisat bölümünde okuyan 40 öğrencinin bir aylık giderleri aşağıda verilmiştir. Sıklık ve birikimli sıklık çizelgelerini 6 sınıf için düzenleyiniz.

100	150	200	200	250	300	300	300	300	300
350	350	350	400	400	400	400	400	400	400
450	500	500	500	500	500	500	500	500	500
500	500	600	600	600	650	750	800	800	850

3) Aşağıda 10 şirketin tahvil fonlarının beş yıldaki getiri yüzdeleri verilmiştir. Getiri yüzdelerinin örneklem ortalamasını, varyansını ve standart sapmasını bulunuz.

99.7	77.5	86.3	85.6	69.5	96.6	83.1	75.1	90.4	82.1

4) Aşağıda 9 hisse senedinin belirli bir gündeki staış fiyatları verilmiştir. Satış fiyatlarının örneklem ortalamasını, ortancasını, varyansını ve standart sapmasını bulunuz.

7.40	8.80	4.04	7.45	6.00	5.10	9.90	3.34	18.70

5) Bir apartmandaki 10 dairenin kira tutarları (YTL) aşağıdaki gibidir. Kira tutarlarının ortalamasını, ortancasını, tepe değerini ve varyansını hesaplayınız.

300	500	350	400	600	450	550	500	350	500

6) Bir tekstil firması yılda 3 milyon triko, 500 bin tişört, 400 bin gömlek ve 500 bin çorap üretmektedir. Sıklık dağılımını oluşturunuz ve daire dilim grafiği çiziniz.

BÖLÜM 2 OLASILIK

2.1. GİRİŞ

Olasılık, günlük yaşamımızda sıkça kullandığımız, yararlandığımız bir kavramdır. Örneğin meteoroloji uzmanı sabah haberlerinde o gün %80 olasılıkla yağmur yağacağını, sağlık uzmanları sigara içenlerin içmeyenlere oranla kansere yakalanma riskinin daha yüksek olacağını, sınavı başarısız geçmiş bir öğrenci o dersten geçme şansının çok az olacağını söyler. Herhangi bir olayın meydana gelme şansını ölçmeyle ilgilenen olasılık, istatistiğin önemli bir bölümünü oluşturmaktadır. İstatistiğin çıkarsama (öngörü) temelini oluşturan olasılık, belirsizlik durumunda sağlıklı kararlar vermeyi sağladığı için, planlama çalışmalarında yoğun bir biçimde kullanılmaktadır. Örneğin bir firmanın gelecek yıldaki satış kestirimleri, bir kısmı gerçekleşecek bir kısmı gerçekleşmeyecek bir çok varsayıma dayalıdır. Bu nedenlerden dolayı olasılık kuramı, bizlere belirsizlik altında ya da mevcut bilgilerin tam ve sağlıklı olmaması gibi durumlarda doğru ve sağlıklı kararlar verebilmede yardımcı olacaktır. Bu bölümde olasılığa ilişkin temel kavramlar verilecek, daha sonra olasılık hesaplama kuralları ele alınacaktır.

Olasılığın Başlangıcı…

Arkeologlar tarih öncesi barınaklarda, medeniyetler boyunca değişik toplumlar için oyun ve şans oyunlarının önemli bir eğlence aracı olduğunu gösteren kanıtlar buldular. Yunan, Mısır, Çin ve Hindu hanedanlarının büyük matematiksel buluşları ve insanların oyun oynamaya olan eğilimleri dikkate alındığında, olasılık matematiğinin diğerleri arasında en erken gelişmeye başladığı tahmin edilebilir. Buna karşın, 17. yy da fransız matematikçiler Pierre de Fermat ve Blaise Pascal'a kadar, olasılık matematiğinde dikkate değer bir gelişme olmamıştır.

Noktalar Problemi

Rönesans Avrupa'sında matematiksel olasılığın gelişimine ilham olan problem budur. Kısaca şöyle tanımlanabilir:

İki eşit yetenekte oyuncu belli bir para karşılığında bir şans oyunu oynarlarken, oyun kesiliyor. Bu aşamadaki oyun skoruna göre para nasıl bölünmelidir?

Burada "eşit yetenekli" her oyuncunun oyuna eşit kazanma şansı ile başladığı anlamına gelir. Aşağıdaki senaryoyu takip ediniz.

Pascal ve Fermat Paris'te bir kafede oturuyorlar ve daha zor senaryolar tartışarak geçen yorucu saatlerden sonra, oyunların en kolayı olan, para atma oyununu oynamaya karar verirler. Eğer tura gelirse Fermat, yazı gelirse Pascal bir puan alacaktır. İlk kez on puan alan kazanacaktır. Sonuçta ne olursa olsun birinin diğerini yemeğe götüreceğini bilerek, her biri ortaya 50 Frank para koyuyor, böylece ortada 100 Frank bahis oluyor. Kazanan hepsini alacak. Ama sonra ilginç birşey oluyor. Fermat 8 e karşı 7 yenerken, bir arkadaşının hasta olduğuna dair bir mesaj alıyor ve Toulouse'a gitmesi gerekiyor. Mesajı getiren adam, hemen hareket etmek şartıyla onları götürebileceğini söylüyor. Tabii ki Pascal durumu anlayışla karşılıyor, ama ortadaki parayı nasıl bölüştürecekler?

Fermat bir mektubunda şu teklifi sunuyor:

"Değerli Blaise,
100 Frankı bölüştürme problemi hakkında, adil olduğunu düşüneceğin bri çözüm buldum. Benim sadece 2 senin de 3 puana ihtiyacın olduğunu düşünürsek, 4 atışta oyun kesinlikle bitmiş olacaktı.

Bu dört atışta, eğer galibiyet için gereken üç puanı alamazsan, bu benim kazanmam için gereken iki puanı aldığım anlamına gelir. Benzer yolla, eğer ben iki puanı alamazsam, sen kazanman için gerekli olan en az üç puanı almış ve kazanmışsın demektir. Böylece, aşağıdaki olası oyun senaryolarının eksiksiz olduğuna inanıyorum. 'h' harfi tura, 'y' harfi de yazıyı göstermektedir. Benim kazandığım senaryolara yıldız koydum.

t t t t *	t t t y *	t t y t *	t t y y *
t y t t *	t y t y *	t y y t *	t y y y
y t t t *	y t t y *	y t y t *	y t y y
y y t t *	y y t y	y y y t	y y y y

Bu çıktıların hepsinin eşit olasılıklı olduğunda hem fikir olduğumuzu düşünüyorum. Bu yüzden, parayı benim lehime olarak 11:5 oranında dağıtmamız gerektiğine inanıyorum, yani ben (11/16)*100=68.75 Frank almalıyım ve sen de 31.25 Frank almalısın tabii.
Paris'te herşeyin yolunda olması temennisiyle,
Arkadaşın ve meslektaşın,
Pierre"

2.2. RASGELE DENEY, RASGELE SONUÇ

Pekçok gözlemden sadece bir tanesinin gerçekleşme sürecine "**Rasgele (Rassal) Deney (Deneme)**" denir. Örneğin; madeni para ile yapılan yazı-tura atışı, zar atışı, üretilen bir malın hatalı olup-olmaması, bir hisse senedinin karlı olup-olmaması, bir öğrencinin dersi geçme notu,...

Rasgele bir deneyin mümkün her bir sonucuna "**rasgele sonuç**" denir. Tüm mümkün sonuçlarının kümesine ise "**örneklem uzayı**" denir ve "S" ile gösterilir.

Örneğin;

RASGELE DENEY 1: Bir madeni paranın bir defa atışı.
RASGELE SONUÇ 1: Yazı ya da Tura
S={Yazı, Tura}

RASGELE DENEY 1: Bir zar atışı
RASGELE SONUÇ 1: Bir, İki, Üç, Dört, Beş ya da Altı
S={1, 2, 3, 4, 5, 6}

RASGELE DENEY 1: Üretilen bir malın hatalı olup-olmaması
RASGELE SONUÇ 1: Hatalı ya da hatasız
S={Hatalı, Hatasız}

RASGELE DENEY 1: Bir öğrencinin dönem başarısı.
RASGELE SONUÇ 1: A1, A2, B1, B2, C, F3, F2, F1
S={Yazı, Tura}

Bir deneyin örneklem uzayı Venn ya da ağaç diyagramı çizilerek de oluşturulabilmektedir. Venn diyagramı, bir deneyin tüm olası sonuçlarının (kare, dikdörtgen ya da daire gibi) bir şekille gösterilmesidir. Ağaç diyagramındaysa her bir sonuç, ağacın bir dalıyla ifade edilmektedir. Venn ve ağaç diyagramları olasılık kavramlarının, görsel ifade yoluyla kolay anlaşılmasına yardımcı olmaktadır.