ANADOLU ÜNİVERSİTESİ

İST 213 OLASILIK DERSİ

TANIMLAR VE VERİ SINIFLAMASI

DOÇ. DR. NİHAL ERGİNEL

TANIMLAR VE VERİ SINIFLAMASI

Olasılık, ilgilenilen olay/olayların meydana gelme olabilirliğinin ölçülmesidir. Olasılık istatistiğin temelidir.

İstatistik, geçmiş ve şimdiki durumdan sonuç çıkarabilmek için verilerin toplanması, analiz edilmesi ve yorumlanması ile gelecek hakkında karar vermeye yardımcı olan bir bilim dalıdır.

- Bir toplumun gelir seviyesi,
- Enflasyon rakamları,
- Nüfus, eğitim istatistikleri,
- Tarım, dış ticaret istatistikleri,

örnek olarak verilebilir.

İstatistiğin oluşunu kavrayabilmek için olay kavramını bilmek gereklidir.

<u>Yığın olay</u>: Birbirine benzemeyen, bazı ortak tarafları olan olaylardır.

- Bir olay bütün olayları temsil edemez. Nüfus, yaş, gelir seviyesi, verimlilik gibi.
- Dolayısıyla aynı koşullar altında farklı sonuçlar ile karşılaşılabilir.
- Bu olaylar, belirli olmayan modeller yardımı ile incelenir.

<u>Tipik olay:</u> Birbirinin tam olarak benzeri olaylardır.

- Deney ne kadar tekrar edilirse edilsin sonuç değişmez.
- Bu olaylardan biri benzer olayları temsil eder. Suyun belli sıcaklıkta kaynaması gibi.
- Aynı koşullar altında aynı sonucu veren olaylar , belirli modeller ile incelenir.

! Olasılık ve istatistik yığın olaylar ve belirli olmayan modellerle ilgilenir.

Olaylar,

- hem genel nedenler
- hem de rassal nedenlerden

etkilenir.

<u>Genel nedenler</u>, aynı gruptaki bütün olaylar üzerinde hep aynı yönde ve aynı derecede etki yaparlar. Irkın boy, renk vb. üzerine etkisi gibi.

<u>Rassal nedenler</u>: bazı olayları aynı, bazı olayları zıt yönde etkiler ve etki dereceleri de farklıdır.

- Hava durumu,
- iklim,
- toprağın cinsi,
- tohum kalitesi,
- tarım tekniği verimliliği etkileyen rassal nedenlerdir.

Bir olay hem genel hem de rassal nedenlerden etkilenebilir. Dolayısıyla aynı koşullar altında farklı sonuçlar ortaya çıkar. Bu olaylar

- "**rassal olay**" veya
- "**rassal deney**"

olarak da adlandırılabilir.

<u>Örnek uzayı:</u>

Rassal deney sonucu karşılaşılan tüm sonuçların oluşturduğu evrensel kümeye **örnek uzayı** (sample space) denir.

<u>Birim (unit):</u>

Her bir rassal olay veya deney sonucuna **birim** denir.

- ❖ Birim sayılabilir veya ölçülebilir olmalıdır.
- ❖ Tanımı açık ve kesin olmalıdır. Örneğin; ailenin geliri;
  - -toplam gelirin birey başına düşen miktarı mı,
  - -toplam geliri mi,
  - -yoksa aile reisinin gelirini mi ifade ettiği açık olmalıdır.

Anakütle (population):

İstatistiksel tekniklerle incelenmek istenen ve bilimsel homojenliğe sahip birimlerin oluşturduğu topluluğa **anakütle** denir.

Örnek (sample):

Anakütleyi oluşturan birimlerin bir kısmını içeren topluluğa **örnek** denir.

FREKANS DAĞILIMLARI

ÖRNEK:

Bir süreçten alınan birimlik örneğin ağırlıkları gr. cinsinden aşağıdaki gibidir. Buna göre, seriyi sınıflandırınız.

60	33	85	52	65	77	84	65	57	74
71	81	35	50	35	64	74	47	68	54
80	41	61	91	55	73	59	53	45	77
41	78	55	48	69	85	67	39	76	60
94	66	98	66	73	42	65	94	89	88

ÇÖZÜM:

2^m Kuralı

$2^1 = 2$

$2^2 = 4$

$2^3 = 8$

$2^4 = 16$

$2^5 = 32$

$2^6 = 64 > n = 50$ (m=6, sınıf sayısı= 6)

ÇÖZÜM:

Sturges kuralı

m = 1+3,22. log n

= 1+3,322. log 50 = 6,64 ➡ 7

h(sınıf aralığı) = $\frac{98-33}{6}$ = 10,83 ➡ 11

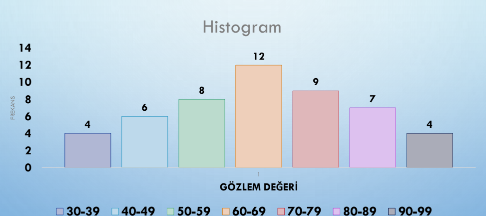
= $\frac{98-33}{7}$ = 9,3 ➡ 10

ÇÖZÜM:

Sınıflar	f_i
30-39	4
40-49	6
50-59	8
60-69	12
70-79	9
80-89	7
90-99	4

60	33	85	52	65	77	84	65	57	74
71	81	35	50	35	64	74	47	68	54
80	41	61	91	55	73	59	53	45	77
41	78	55	48	69	85	67	39	76	60
94	66	98	66	73	42	65	94	89	88

Grafik gösterim:

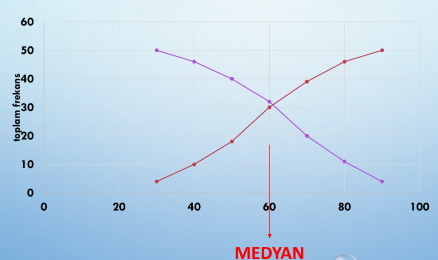


BİRİKİMLİ FREKANS DAĞILIMLARI

Sınıflar	f_i
30-39 ($30 \leq x < 40$)	4
40-49 ($40 \leq x < 50$)	6
50-59 ($50 \leq x < 60$)	8
60-69 ($60 \leq x < 70$)	12
70-79 ($70 \leq x < 80$)	9
80-89 ($80 \leq x < 90$)	7
90-99 ($90 \leq x < 100$)	4

-den çok	
Sınıflar	f_i
30' dan çok	50
40' dan çok	46
50' den çok	40
60' dan çok	32
70' den çok	20
80' den çok	11
90' dan çok	4

-den az	
Sınıflar	f_i
39' dan az	4
49' dan az	10
59' dan az	18
69' dan az	30
79' dan az	39
99' dan az	46
99' dan az	50



TANIMLAYICI (YER) ÖLÇÜLERİ

Aritmetik Ortalama:

Basit aritmetik ortalama:

x_i = i. gözlem değeri

n: veri sayısı

$\bar{x}$ = aritmetik ortalama

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

Frekans Serilerinde Aritmetik Ortalama:

f_i = i. gözlemin tekrar sayısı(frekans)

k = tekrarlanmayan gözlem sayısı

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + .. + f_k \cdot x_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum_{i=1}^{k} f_i \cdot x_i}{\sum_{i=1}^{k} f_i}$$

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

ÖRNEK:

Veriler	f_i
50	3
55	5
60	8
65	10
70	4

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{50*3+5*55+\cdots+70*4}{3+5+8+10+4} = 61,17$$

Sınıflandırılmış Serilerinde Aritmetik Ortalama:

x_i = i. sınıfın sınıf ortalaması

f_i = i. sınıftaki gözlem sayısı

m = sınıf sayısı

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + .. + f_m \cdot x_m}{f_1 + f_2 + \cdots + f_m} = \frac{\sum_{i=1}^{m} f_i \cdot x_i}{\sum_{i=1}^{m} f_i}$$

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

ÖRNEK

Sınıflar	f_i	x_i
(30 ≤ x< 40)	4	34
(40 ≤ x< 50)	6	44
(50 ≤ x< 60)	8	54
(60 ≤ x< 70)	12	64
(70 ≤ x< 80)	9	74
(80 ≤ x< 90)	7	84
(90 ≤ x< 100)	4	94

$$\bar{x} = \frac{\sum_{i=1}^{m} f_i \cdot x_i}{\sum_{i=1}^{m} f_i} = \frac{34*4+44*6+\cdots.....+94*4}{4+6+\cdots.+4} = 64{,}6$$

Ağırlıklı (Tartılı) Aritmetik Ortalama:

w_i = i. gözlemin ağırlığı

x_i = i. gözlemin değeri

t = ağırlıklandırma sayısı

$$\bar{x} = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + .. + w_m \cdot x_m}{w_1 + w_2 + \cdots + w_m} = \frac{\sum_{i=1}^{t} w_i \cdot x_i}{\sum_{i=1}^{t} w_i}$$

$$\sum_{i=1}^{t} w_i = \text{toplam ağırlık}$$

Aritmetik Ortalamanın Bazı Özelikleri

1. Gözlemlerin ortalamadan sapmalarının cebirsel toplamı '0' dır.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n.\,\bar{x}$$

$$= \sum_{i=1}^{n} x_i - \cancel{n}.\,\frac{\sum x_i}{\cancel{n}} = 0$$

2. Aritmetik ortalama gözlemlerdeki aşırı büyük ya da aşırı küçük değerlerden etkilenir.

Aritmetik Ortalamanın Bazı Özelikleri

3. Gözlem değerlerinin belirli bir a sayısından sapmalarının kareleri toplamı ancak ve ancak a = $\bar{x}$ için en küçük olur.

$$\sum_{i=1}^{n}(x_i - a)^2 \text{ , en küçük } \longrightarrow \sum_{i=1}^{n}(x_i - \bar{x})^2$$

4. A sabit bir sayı iken gözlem değerlerine A'nın eklenmesi/çıkarılması durumunda, aritmetik ortalamaya da A eklenir/ çıkarılır.

$$d_i \longrightarrow x_i \pm A \text{ ise, } \bar{d} = \bar{x} \pm A$$

GEOMETRİK ORTALAMA

Aldığımız gözlem değerleri,

- bir önceki gözlem değerine bağlıysa,
- ona göre değişiyorsa ve
- değişimin hızı belirlenmek isteniyorsa

geometrik ortalama kullanılır.

Örneğin; nüfus, millî gelir, bileşik faiz vb.

GEOMETRİK ORTALAMA

x_i = i. gözlemin değeri

n = gözlem sayısı

$\overline{X_G} = \sqrt[n]{x_1 . x_2 \dots \dots . . x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i} = (x_1 . x_2 \dots \dots . . x_n)^{\frac{1}{n}}$

$\log \ \overline{X_G} = \frac{1}{n} \log (x_1 . x_2 \dots \dots . . x_n)$

$= \frac{1}{n} (\log x_1 + \dots\dots + \log x_n)$

$= \frac{1}{n} \sum_{i=1}^{n} \log x_i$

GEOMETRİK ORTALAMA

Frekans serisinde;

$\overline{X_G} = \sqrt[n]{x_1^{f_1} . x_2^{f_2} . x_3^{f_3} \dots . . x_k^{f_k}}$, $\sum_{i=1}^{k} f_i = n$

$\log \ \overline{X_G} = \frac{1}{n} (f_1 . \log x_1 + \dots\dots + f_n . \log x_n)$

$= \frac{1}{n} \sum_{i=1}^{k} f_i . \log x_i$

HARMONİK ORTALAMA

x_i = i. gözlemin değeri

n = gözlem sayısı

$\overline{X_H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$,

$\frac{1}{\overline{X_H}} = \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right) . \frac{1}{n}$

HARMONİK ORTALAMA

Frekans serisinde;
$$\frac{1}{\overline{X_H}} = \frac{1}{n}.\left(\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}\right)$$
$$= \frac{1}{n}\sum_{i=1}^{k}\frac{f_i}{x_i}$$

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

Doğrudan doğruya ifade ettiği anlam yerine, tersine çevrildiğinde taşıyacağı anlama önem verilen, oran vb. niceliklerin ortalamasını bulmak için kullanılır.

Fiyat artışı, verimlilik, hız vb.

KARELİ ORTALAMA

$$\overline{X_K} = \sqrt{\frac{{x_1}^2+{x_2}^2+..+{x_n}^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} {x_i}^2}{n}}$$

Frekans serisinde;

$$\overline{X_K} = \sqrt{\frac{f_1.{x_1}^2+f_2.{x_2}^2+..+f_n.{x_n}^2}{f_1+f_2+..+f_n}} = \sqrt{\frac{\sum_{i=1}^{k} f_i.{x_i}^2}{\sum_{i=1}^{k} f_i}}$$

Diğer ortalamaların hesaplanamadığı durumlarda kullanılır.

MEDYAN

Değerleri küçükten büyüğe sıraladığımızda, grubu iki eşit kısma ayıran değere **medyan** denir.

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & ; \quad n\ tek \\ \dfrac{x_{n/2}+x_{(n+1)/2}}{2} & ; n\ çift \end{cases}$$

- **Frekans serisinde**; birikimli frekans alınır. $\frac{n}{2}$ hangi x_i değerine karşılık geliyor ise, o değer medyandır.

- **Sınıflandırılmış serilerde**; $\cong L_m + \left(\frac{\frac{(n+1)}{2} - T}{f_m}\right)$.h

f_m = medyan sınıfının frekansı
h = sınıfın aralığı
n= örnek sayısı
L_m = medyanı içeren sınıfın alt sınırı
T= medyan sınıfından önceki frekansların toplamı

Medyanın Özellikleri

1-) Seride aşırı büyük ya da aşırı küçük değerler olduğunda merkezi eğilimi aritmetik ortalamadan daha iyi belirler.

2-) Sınıflandırılmış seride açık sınıflar olduğunda, medyan aritmetik ortalamaya üstünlük sağlar.

3-) Hassas değildir, serideki aşırı küçük ya da büyük değerlerden etkilenmez.

4-) Gözlemlerin medyandan sapmalarının mutlak toplamı enküçüktür. $\sum_{i=1}^{n}(x_i - \tilde{x}) \Rightarrow$ enküçük

MOD

Bir seride en çok tekrar eden değere **mod** denir.

- **Frekans serilerinde** en yüksek frekansa sahip değer mod olur.
- **Sınıflandırılmış serilerde:**

$x_{mod} = L_{MO_{alt}} + \frac{a}{a+b}$. h

$x_{mod} = L_{MO_{üst}} - \frac{b}{a+b}$. H

$L_{MO_{alt}}$: mod sınıfının alt sınırı
$L_{MO_{üst}}$: mod sınıfının üst sınırı
a: mod sınıfının frekansı ile ondan bir önceki sınıfın frekansı arasındaki mutlak fark
b: mod sınıfının frekansı ile ondan bir sonraki sınıfın frekansı arasındaki mutlak fark
h: sınıf aralığı

Modun Özellikleri

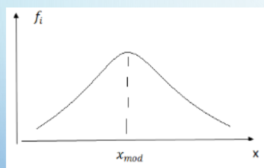
1-) Ortalamalar arasında seriyi en iyi temsil eden değerdir.

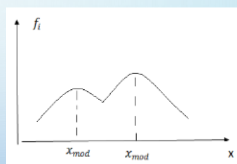
2-) Hesaplanması ve kestirimi kolay olduğundan yaygın bir kullanıma sahiptir.

3-) Aşırı küçük/büyük değerlerden etkilenmez.

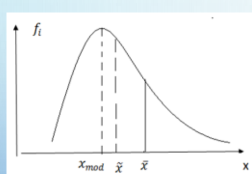
4-) Serinin tüm değerlerine bağlı olmadığından matematiksel işlemlere uygun değildir.

5-) Serinin az olması durumunda veya çok maksimumlu serilerde mod tanımlanamaz.

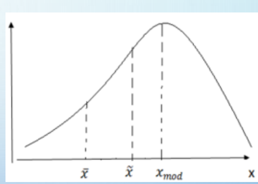




Simetrik serilerde $\bar{x}$ = $\tilde{x}$ = x_{mod}



Pozitif veya sağdan çarpık seri



Negatif veya soldan çarpık seri