IST2084/ IST104.1/ IST104.2
Biyoistatistik
3. Hafta

Doç. Dr. Fatih KIZILASLAN

http://mimoza.marmara.edu.tr/~fatih.kizilaslan/

- **Histogram:** Koordinat eksenleri üzerinde her sınıf için çizilen dikdörtgenlerden oluşan grafiktir. Bu dikdörtgenlerin taban kenar uzunlukları sınıf aralığına eşittir. Diğer kenar uzunluğu da sınıfların frekansına eşittir.

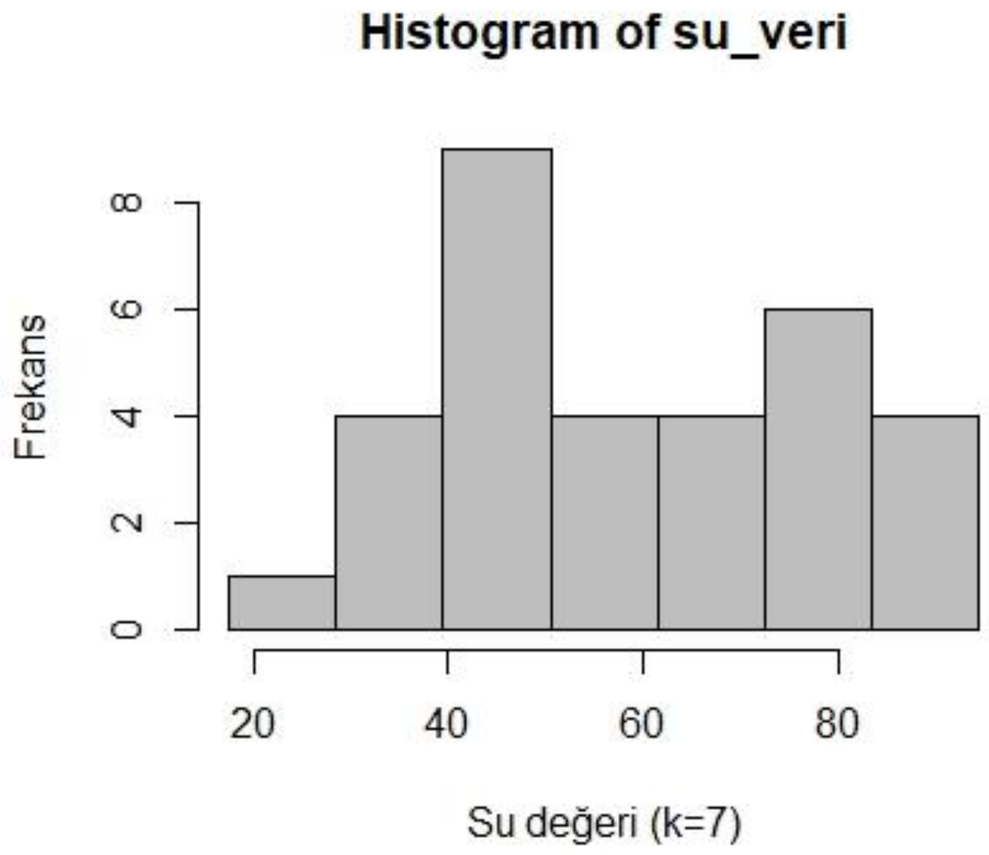
- **Diyagram (çizgi diyagramı veya dağılım poligonu):** Sınıf orta değerleri (sınıfın alt ve üst sınırlarının ortalaması) x ekseni üzerinde frekanslar ise y ekseni üzerinde olmak üzere sınıf orta değerlerinin frekanslarla birleştiği noktaların kesik çizgilerle birleştirilmesiyle oluşturulan grafiktir.

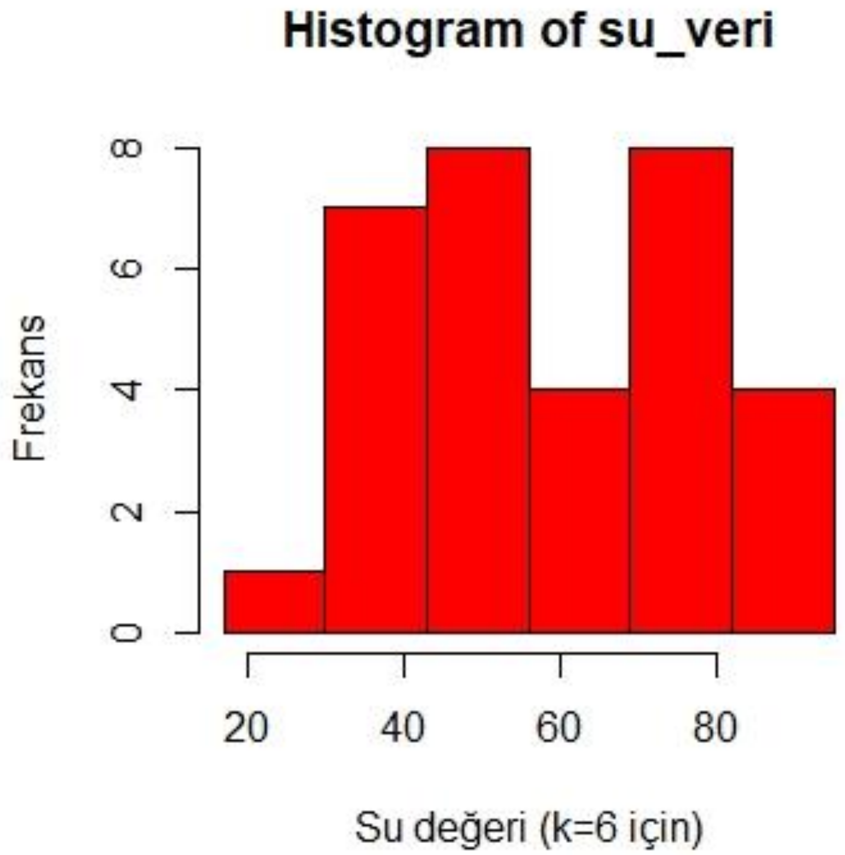
Örnek 3: 32 tür peynirin içerdiği su değerleri (gr/100 gr) aşağıda verilmiştir.

18	34	36	37	39	40	41	41
44	45	46	46	47	49	51	53
57	58	62	65	70	72	73	77
78	79	80	82	84	84	85	94

a) Bu verileri sınıf sayısını 7 alarak frekans tablosunu oluşturunuz. Histogram ve diyagram grafiklerini çiziniz.

b) Bu verileri sınıf sayısını 6 alarak frekans tablosunu oluşturunuz. Histogram ve diyagram grafiklerini çiziniz.


Histogram of su_veri
Frekans
0
2
4
6
8
20
40
60
80
Su değeri (k=7)


Histogram of su_veri
Frekans
0
2
4
6
8
20
40
60
80
Su değeri (k=6 için)

3. Bölüm: Merkezi Eğilim ve Dağılım Ölçüleri

İstatistiksel verideki gözlem değerlerinin etrafında toplandığı değerler **merkezi eğilim ölçüleri** ve birbirlerine göre konumlarını, birbirlerine göre yakınlık ve uzaklıklarını yansıtan değerler de **merkezi dağılım ölçüleri** olarak adlandırılır.

Merkezi eğilim ölçüleri: Aritmetik ortalama, Medyan (ortanca), Mod (tepe değer), Geometrik ortalama, Harmonik ortalama

Merkezi dağılım ölçüleri: Değişim aralığı (range), Varyans ve Standart sapma, Standart hata, Değişim katsayısı

Dağılımın şekli ile ilgili olarak **çarpıklık (skewness)** ve **basıklık (kurtosis)** katsayılarını inceleyeceğiz.

Aritmetik Ortalama

- Gözlem değerlerinin toplamının gözlem sayısına bölümü olarak tanımlanır.
- Gözlem değerlerinin etrafında toplandığı merkezi ifade eder.
- En çok kullanılan merkezi eğilim ölçüsüdür.
- Popülasyon (kitle) için aritmetik ortalamayı μ (Mu) ile gösteririz.
- Örneklem için aritmetik ortalamayı $\bar{x}$ ile gösteririz.
- Sadece nicel veriler için mevcuttur.

- **Sınıflandırılmamış veri için aritmetik ortalama**

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

x_i: i. gözlem değeri, n: toplam gözlem sayısı

- **Sınıflandırılmış veri için aritmetik ortalama**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i \, s_i$$

f_i: i. sınıfın frekans değeri, s_i: i. sınıfın orta değeri, k: sınıf sayısı

Medyan (Ortanca)

- Bir verideki gözlem değerleri küçükten büyüğe doğru sıralandığında ortadaki değere medyan denir.
- Medyan veriyi ortadan ikiye böler.
- Nicel ve sıralama ölçeğine göre ölçülen değişkenler için kullanılabilir.
- Aşırı uç değerlerden (aykırı değer) etkilenmez. Veride aykırı değerler olması durumunda ortalama yerine tercih edilir.

- **Sınıflandırılmamış veri için medyan**

Toplam gözlem sayısının çift veya tek olmasına göre aşağıdaki gibi bulunur.

$$x_{Medyan} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n\ tek \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & n\ çift \end{cases}$$

$x_{(i)}$: Gözlemler küçükten büyüğe sıralandığında i. gözlem değeri

Örnek: 15,16,18,14,12,17,18,20,19,14,15,18 için medyanı bulalım.

12 gözlem vardır. Küçükten büyüğe sıralayalım.

12,14,14,15,15,16,17,18,18,18,19,20

$$x_{Medyan} = \frac{x_{(6)} + x_{(7)}}{2} = \frac{16+17}{2} = 16.5$$

Mod (Tepe değer)

- Bir veride en çok tekrarlana değere mod (tepe değer) denir.
- Nitel ve nicel veriler için uygundur.
- Gözlemlerin tümü işleme katılmadığından uç değerlerden etkilenmez.
- Her gözlem bir kez ortaya çıkmış ise mod mevcut değildir.
- Bir veride birden fazla mod değeri olabilir.
- **Sınıflandırılmamış veri için mod:** Veride en çok tekrarlanan gözlemdir.

Örneğin 12,14,14,15,15,16,17,**18,18,18**,19,20 için mod değeri 18 dir.

Geometrik Ortalama

- Gözlem sonuçlarında mutlak farklar yerine oransal farklar ile ilgileniliyorsa yani gözlemlerin her biri önceki gözlem değerine bağlı olarak oluşuyorsa bu durumda geometrik ortalamayı kullanmak uygundur.

$$G(Geometrik\ Ortalama) = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

- **Örnek:** Bir şehirdeki son dört yıllık nüfus değerleri (bin olarak) 100,180,210,300. Bu şehrin son dört yıllık ortalama nüfus artışı yüzde kaçtır?
- **Çözüm:** Yıllık artış oranları 180/100=1.8, 210/180=1.16, 300/210=1.42

$$G = \sqrt[3]{1.8\ 1.16\ 1.42} = 1.437$$

Böylece son dört yıllık ortalama nüfus artışı %43.27 dir.

Harmonik Ortalama

- Ekonomik olaylarda 1 birim ile alınan ortalama miktara veya bir ürünün bir biriminin üretimi için harcana ortalamaya gereksinim olduğunda harmonik ortalama kullanılır.

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

biçiminde hesaplanır.

Örnek: Bir tekstil fabrikasında çalışan dört kişinin bir pantolonu ütüleme süreleri 10dk, 6dk, 4dk, 5dk dır. Buna göre bu fabrikada bir pantolon ortalama kaç dakikada ütülenir?

$$H = \frac{1}{\frac{1}{4}\left(\frac{1}{10} + \frac{1}{6} + \frac{1}{4} + \frac{1}{5}\right)} = 5.58\ dk$$