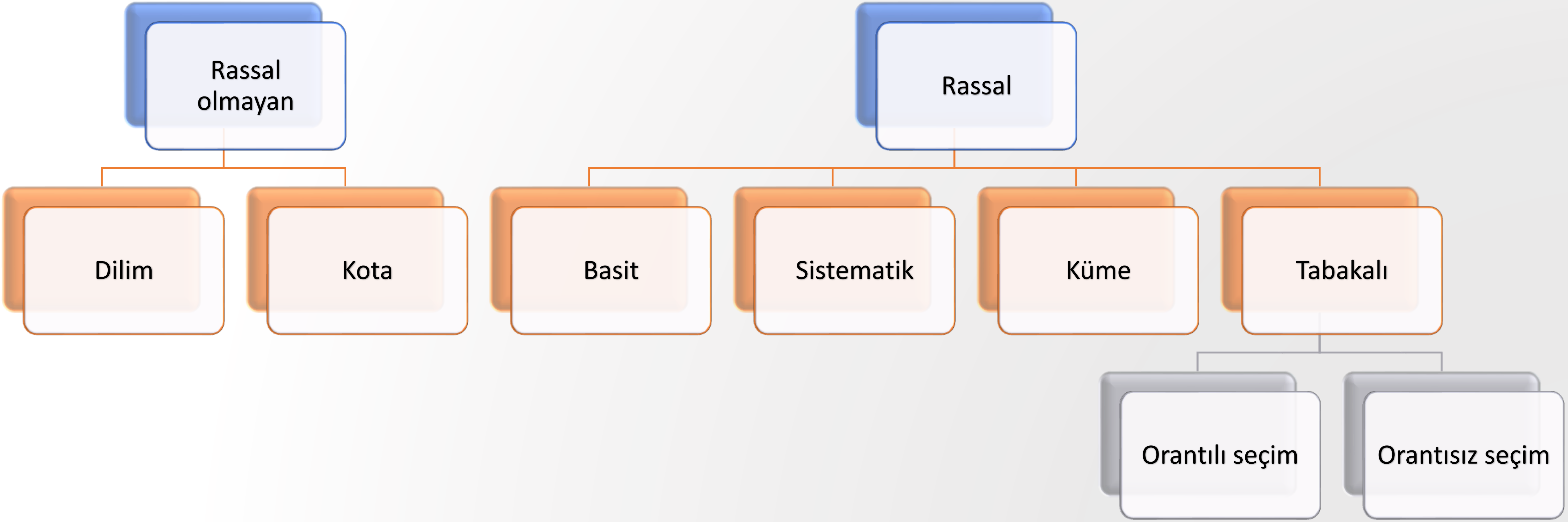
İstatistik çalışmaları

Oya Şanlı

Ana kütledeki her birimin örnek kütleye girme olasılığına göre



Verilerin toplanması - Seriler

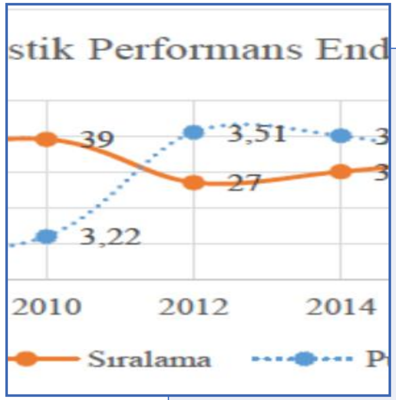
- Zaman ve Mekan(2 sütunlu)
  - zaman
  - mekan
- Nitel(kalitatif-2 sütunlu)
- Nicel
  - Basit (tek sütunlu
  - sınıflandırılmış(2 sütunlu)
  - Gruplandırılmış(2 sütunlu)

Grup aralığı

- Grup sayısı K= 1 + 3.3log(n) (n= örneklenen veri sayısı- ana kütle)
- Değişim genişliği DG= X_{enb} – X_{enk} (örneklenen verideki en büyük ve en küçük farkı)

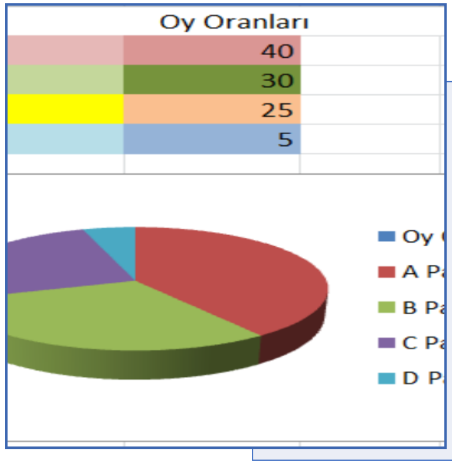
Grup aralığı GA = DG/K

Verilerin sunulması



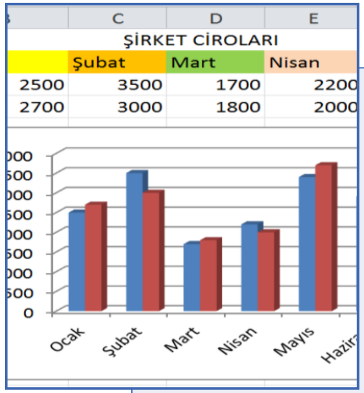
Zaman serileri

- X ekseni zaman
- Y ekseni örneklenen birimler
- KARTEZYEN GRAFİK



Nitel seriler

- PASTA GRAFİĞİ



Nicel seriler

- X ekseni sınıflandırılmış seri değerleri
- Y ekseni frekanslar
- SÜTUN/ÇUBUK GRAFİK

Merkezi eğilim ölçüleri

- Bir örnekleme sonucunda toplanan verilerin hangi değer etrafında toplandığını gösteren ve verilerin oluşturduğu seriyi temsil eden rakama “ortalama” denilir.
- Örneklenen verilerin herhangi birinde meydana gelecek değer değişikliği, ortalamanın değerinde de değişikliğe yol açıyorsa, ortalamanın hesaplanmasında tüm örneklenen verileri dikkate alan duyarlı ortalama, bu değişim ortalamanın değerini etkilemiyorsa, verilerin tümünü dikkate almayan duyarsız ortalama yöntemleri kullanılabilir.

Merkezi eğilim ölçüleri

Normal bir seride ortalamalar arasında aşağıdaki gibi bir büyüklük ilişkisi vardır.

Kareli > Aritmetik > Geometrik > Harmonik Ortalama

K > X > G > H

Duyarlı ortalamalar

- Basit aritmetik ortalama
- Ağırlıklı Aritmetik ortalama
- Geometrik ortalama
- Harmonik ortalama
- Kareli ortalama

Duyarsız ortalamalar

- Kartil ortalama
- Mod
- Medyan

Aritmetik ortalama

Nicel örnekleme verileri toplamının veri sayısına oranı

$$\bar{X} = \frac{\sum X_i}{n}$$

Basit serilerde

$$\bar{X} = \frac{\sum X_i . f_i}{\sum f_i}$$

Sınıflandırılmış serilerde

$$\bar{X} = \frac{\sum m_i . f_i}{\sum f_i}$$

Gruplandırılmış serilerde

Küçültülmüş değerlerle aritmetik ortalama

Aritmetik ortalamanın hesaplanmasında, verilerin çok fazla ve büyük değerlerden oluşması durumunda, aritmetik ortalamaya en yakın olduğu düşünülen bir grup ortalaması (genellikle en büyük frekansa sahip grubun ortalaması), geçici ortalama olarak seçilir (m0).

$$\bar{X} = m_0 + (\bar{u} \, . \, A)$$

(A Grup ortalamalarının grup aralığı)

(u_i küçültülmüş değerlerden oluşan grup ortalamaları)

$$u_i = \frac{m_i - m_0}{A}$$

($\bar{u}$ Küçültülmüş değerlerden oluşan serisinin aritmetik ortalaması)

$$\bar{u} = \frac{\sum u_i . f_i}{\sum f_i}$$

Aritmetik ortalamanın bazı matematiksel özellikleri

Aritmetik ortalama, serideki aşırı değerlerden oldukça fazla etkilenen hassas bir ortalamadır.

$$X_i = Y_i + Z_i \text{ ise } \bar{X} = \bar{Y} + \bar{Z} \text{ dir.}$$

$$\sum\left(X_i - \bar{X}\right)^2 = \text{En küçük}$$

$$\sum\left(X_i - \bar{X}\right) = 0$$

Ağırlıklı Ortalama

- Örneklenen verilerin önem derecesinin farklı olduğu durumlarda ise, bu önem derecelerinin de hesaplamalara dahil edilmesi gerekir.

$$\overline{X} = \frac{\sum a_i.X_i}{\sum a_i}$$

Basit seriler

$$\overline{X} = \frac{\sum a_i.X_i.f_i}{\sum a_i.f_i}$$

Sınıflandırılmış seriler

$$\overline{X} = \frac{\sum a_i.m_i.f_i}{\sum a_i.f_i}$$

Gruplandırılmış seriler

a_i = i'inci örneğin önem veya ağırlığıdır

Geometrik ortalama

Verilerin ani ve anormal farklılıklar gösterdiği ya da yüzde/oranlarla ifade edildiği durumlarda geometrik ortalama kullanılır

Veri değerlerinin herhangi birinin sıfır veya sıfırdan küçük olması durumunda geometrik ortalamanın hesaplanması mümkün olamamaktadır.

İki farklı şekilde hesaplanır

- Bir veri setinde bulunan n adet birimin çarpımının n'inci dereceden kökünün alınmasıyla elde edilen değer

$$G = \sqrt[n]{X_1.X_2.X_3.X_4.......X_n}$$

- Verilerin logaritmaları alınarak bulunacak logaritmik aritmetik ortalamanın eksponansiyeli (anti logaritması) hesaplanarak geometrik ortalama elde edilir.
  - Basit serilerde

$$\ln G = \frac{\sum \ln X_i}{n}$$

  - Sınıflandırışmış serilerde

$$\ln G = \frac{\sum \ln X_i.f_i}{\sum f_i}$$

  - Gruplandırılmış serilerde

$$\ln G = \frac{\sum (\ln m_i) f_i}{\sum f_i}$$

n'inci dereceden kök alma zorluğu nedeniyle logaritmalar alınarak hesaplama tercih edilmektedir.

Harmonik ortalama

- Bir serideki gözlem değerlerinin terslerinin aritmetik ortalamasının tersidir.

Oran, yüzde ve bölme şeklinde ifade edilen seri değerlerinin ortalamasını hesaplamada aritmetik ortalama uygun bir ortalama olmayıp, bu gibi durumlarda harmonik ortalamanın kullanımı tercih edilir

Basit serilerde

$$H = \frac{n}{\sum \frac{1}{X_i}}$$

Gruplanmış serilerde

$$H = \frac{\sum f_i}{\sum \frac{f_i}{m_i}}$$

Sınıflandırılmış serilerde

$$H = \frac{\sum f_i}{\sum \frac{f_i}{X_i}}$$

Bir seri değerinden aritmetik ortalamanın çıkarılması ile elde edilen sapmalar serisinin toplamı daima sıfıra eşittir.

Bu nedenle, sapmalar serisinin ortalamasının hesaplanmasında da kareli ortalama kullanılır.

$$K = \sqrt{\frac{\sum X_i^2}{n}}$$

Basit serilerde

$$K = \sqrt{\frac{\sum f_i . X_i^2}{\sum f_i}}$$

Sınıflandırılmış serilerde

$$K = \sqrt{\frac{\sum f_i . m_i^2}{\sum f_i}}$$

Gruplandırılmış serilerde

Kareli ortalama

- Serideki değerlerin karelerinin aritmetik ortalamasının kareköküdür
- Genellikle bir seride sıfırdan küçük terimler bulunduğu veya terimler toplamı sıfıra eşit olduğunda kullanılmaktadır.

Medyan

Serideki değerler küçükten büyüğe sıralandığında tam ortaya düşen ve seriyi iki eşit parçaya bölen değere medyan (ortanca) denilir.
Örneklenen verilerin dağılımının çarpık olduğu veya seride aşırı küçük/büyük değerlerin bulunduğu durumlarda, merkezi eğilim ölçüsü olarak medyanın kullanımı tercih edilebilmektedir.

Basit seriler

- Önce veriler küçükten büyüğe sıralanırlar, sonra serideki veri sayısının tek veya çift sayıda olmasına göre medyan hesaplanır.

Sınıflandırılmış seriler

- Frekanslar, küçükten büyüğe kümülatif (toplam) frekanslar haline dönüştürülür.

Gruplandırılmış seriler

- Medyan grubunun belirlenerek medyan hesaplanır.

Medyan

Basit seriler

- Veri sayısı tek ise $\frac{n+1}{2}$
- Veri sayısı çift ise $\frac{n}{2}$ ve $\frac{n}{2}+1$ gözlem değerlerinin aritmetik ortalaması

Sınıflandırılmış seriler

- $\left(\left(\sum f_i\right)+1\right)/2$ inci frekansı içeren terimdir

Gruplandırılmış seriler

$$M_e = L_m + \left\{ S_m \cdot \frac{\left(\frac{\sum f_i}{2}\right) - \left(\sum_{i=1}^{m-1} f_i\right)}{f_m} \right\}$$

L_m = medyan grubunun alt sınırı

S_m = medyan grubunun aralığı

$\left(\sum f_i / 2\right)$ = toplam frekansın yarısı

$\left(\sum_{i=1}^{m-1} f_i\right)$ = medyan grubundan bir önceki grubun toplam frekansı

f_m = medyan grubunun frekansı

Mod

- Bir seride en çok tekrarlanan değere veya frekansı en büyük olan değere mod denir.
  - Anormal aşırı değerlerin etkisi altında kalmaz.
  - Tüm veri değerlerini göz önünde bulundurmadığı için tutarlı olmayan bir merkezi eğilim ölçüsüdür.
  - Sınıflandırılmış ve gruplandırılmış seriler için hesaplanabilir.

Mod

Sınıflandırılmış seriler

En yüksek frekansa karşılık gelen değer

Gruplandırılmış seriler

L_{mo} = mod grubunun alt sınırı
S_{mo} = mod grubunun aralığı
f_{mo} = mod grubunun frekansı
f_{mo-1} = mod grubundan önceki grubun frekansı
f_{mo+1} = mod grubundan bir sonraki grubun frekansı

$$M_o = L_{mo} + \left\{ S_{mo} \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} \right\} \quad \Delta_1 = f_{mo} - f_{mo-1} \quad \Delta_2 = f_{mo} - f_{mo+1}$$

Dağılım ölçüleri

- Ortalamalar, rassal örneklenmiş verilerin merkezi eğilim ölçülerini göstermekle birlikte, bu değer çevresindeki yayılımın büyüklüğü hakkında bir bilgi vermez.
- Dağılım ölçüsü, seriyi oluşturan verilere sabit bir sayı eklendiğinde veya çıkarıldığında değeri değişmeyen ölçüdür.

Dağılım ölçüleri

Değişkenlik aralığı

Veri içindeki en küçük ve en büyük değer arasındaki fark

$R = X_{enb} - X_{enk}$

Veri sayısı çok olduğunda güvenilir değildir

Varyans ve standart sapma

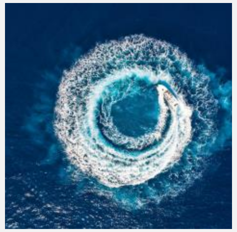
Veri dağılımı veya yayılımının büyüklüğünü ölçmek için en çok kullanılan parametre

Basit seriler

Sınıflandırılmış seriler

Gruplandırılmış seriler

Değişkenlik katsayısı



Ortalamaya göre yayılımın büyüklüğünü gösteren bir katsayı

$$C = \sigma / \bar{X}$$

Aritmetik ortalama ve standart sapmanın farklı sonlandığı durumlarda, mutlaka değişkenlik katsayısını da hesaplayarak karar vermek gerekir

Varyans ve Standart sapma

Basit seriler

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n}$$

S_2 = dağılımın varyansı,
X_i = i'inci rassal örneklenmiş değişkenin değeri
$\bar{X}$ = dağılımın örnek kütle aritmetik ortalaması
n = kütlenin örnek sayısı

Basit serilerde eğer n<30 ise, varyans hesaplamada payda "n-1" olur.

Sınıflandırılmış seriler

$$S^2 = \frac{\sum_{i=1}^{k} \left(X_i - \bar{X}\right)^2 . f_i}{\sum_{i=1}^{k} f_i}$$

k = sınıf sayısı
f_i = i'inci sınıfın frekansı

Gruplandırılmış seriler

$$S^2 = \frac{\sum_{i=1}^{l} \left(m_i - \bar{X}\right)^2 . f_i}{\sum_{i=1}^{l} f_i}$$

l = grup sayısını
m_i = i'inci grubun ortalamasını
f_i = i'inci grubun frekansı

Varyansın büyük olması, değişkenin ortalama çevresindeki yayılımının büyük olduğunu gösterir, yani ortalamadan uzak değerler alma olasılığının daha büyük olduğunu söylemek mümkündür