İstatistik I

4. Hafta

28 Ekim 2021

4. Merkezi Değişim Ölçüleri

www.umitsarp.com

Ümit SARP, umit.sarp@ikcu.edu.tr

Giriş

- Merkezi Değişim Ölçüleri
  - Sapma
  - Varyans
  - Standart sapma
  - Değişim katsayısı
  - Chebyshev teoremi

Merkezi Değişim Ölçüleri

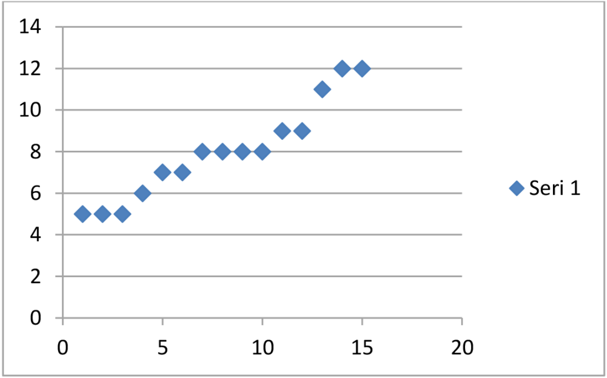
Bir serideki gözlemlerin (birimlerin) birbirinden ya da herhangi bir ortalama değerden uzaklıklarının çeşitli ölçümlerine Merkezi Değişim Ölçüleri denir. Bir seriyi özetlemede ortalamalar tek başına yeterli bir veri değildirler. Ortalamaların yanısıra değişim ölçülerine de ihtiyaç vardır [3].

Örneğin aşağıda ortalamaları birbiriyle aynı ama dağılımları farklı üç seriyi inceleyelim:

Merkezi Değişim Ölçüleri

Seri1: 5 5 5 6 7 7 8 8 8 8 9 9 11 12 12

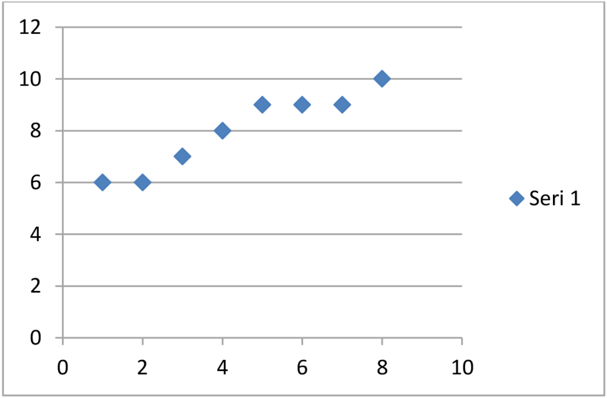
$\bar{X}_1 = 8$



Merkezi Değişim Ölçüleri

Seri2: 6 6 7 8 9 9 9 10

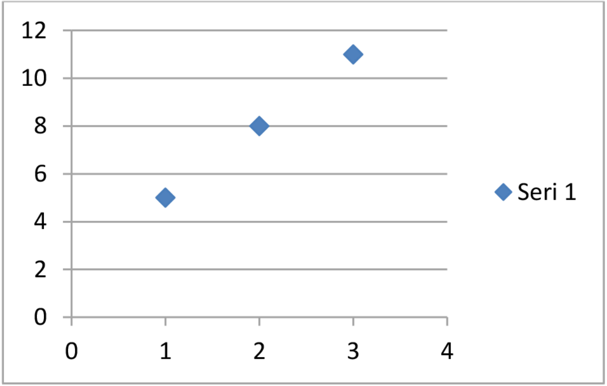
$$\bar{X}_2 = 8$$



Merkezi Değişim Ölçüleri

Seri3: 5 8 11

$\bar{X}_3 = 8$



Merkezi Değişim Ölçüleri

Her üçünün de ortalaması birbiriyle aynı olmasına rağmen, veri sayıları, değişimleri, dağılımları birbirinden farklıdır. Öyleyse artık değişim ölçülerinin neler olduklarına ve nasıl hesaplandıklarına ilişkin açıklamalarımıza geçebiliriz.

Merkezi Değişim Ölçüleri

Verilerin dağılımını ölçmede, gözlem değerlerinin ortalamadan sapmalarını kullanabiliriz.

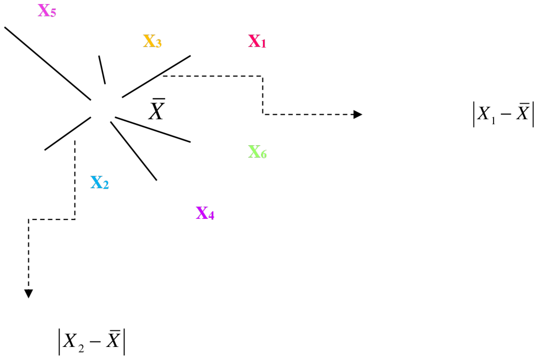
Bu sapmalar yani

$$\left|X_i - \bar{X}\right|$$

ne kadar büyükse, X_i gözlemi ortalamadan o kadar uzakta demektir.

Merkezi Değişim Ölçüleri

Birimlerin Ortalamadan Uzaklıkları (sapmaları)



Sizce bu sapmalar ne anlama geliyor ve nasıl ölçülecek?

Merkezi Değişim Ölçüleri

Mesela, $\bar{X}$ 'dan sapmalar ile ölçülebilir mi?

Hayır ölçülemez, çünkü sapmaların genel toplamı daima sıfırdır.

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

Merkezi Değişim Ölçüleri

Varyans

Öyleyse biz de sapmaların karelerini alıp işe başlarız yani aritmetik ortalama değil de, kareli ortalamayı kullanırız. Sapmaların kareli ortalaması bulunduğunda, istatistik teorisinde ve uygulamalarda en yaygın kullanılan dağılma ölçüsü olan standart sapmayı yani σ ya da s'yı elde ederiz. Standart sapmanın karesine yani σ^2 ya da s^2'ye ise varyans denir. Varyans değişimin ölçüsüdür.

σ^2 kitle varyansıdır, s^2 ise örneklem varyansıdır.

Merkezi Değişim Ölçüleri

Standart Sapma

Bir seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan sapmalarının kareli ortalaması standart sapma olarak tanımlanır. Değişkenlik ölçüleri içinde en yaygın ve etkin kullanıma sahip olan standart sapmadır. Standart sapma serideki gözlem değerlerinin tümüne dayanan, dolayısıyla matematiksel bakımdan en kuvvetli değişkenlik ölçüsüdür.

Merkezi Değişim Ölçüleri

Standart Sapma

Bir seriyi oluşturan gözlem değerlerinin birbirine yakınlığı standart sapma ile ifade edilir. Genel olarak seriyi oluşturan gözlem değerleri, hesaplanan ortalamaya yakın değerlerden oluşuyorsa standart sapma değeri küçük, uzak değerlerden oluşuyorsa standart sapma değeri büyük olur.

Standart sapma genellikle örnek kütle için s-(küçük-s) ile evren için σ-(küçük σ) ile gösterilir.

Merkezi Değişim Ölçüleri

Standart Sapma

Dizilerde standart sapma hesaplanmasında;

- Öncelikle serini aritmetik ortalaması ($\bar{X}$) hesaplanır.

Merkezi Değişim Ölçüleri

Standart Sapma

- Hesaplanan aritmetik ortalama, her bir gözlem değerinden çıkarılır ve bu farkların karelerinin toplamıalınır.

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

Merkezi Değişim Ölçüleri

Standart Sapma

- İzleyen formüller uygulanarak standart sapma ve varyans hesaplanır.

Standart Sapma;

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

Varyans;

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

Merkezi Değişim Ölçüleri

Bu formüllerdeki standart sapma ve varyans örnek içindir. Evren için standart sapma ve varyans formülleri yazılırken $\bar{X}$ yerine evren aritmetik ortalaması μ ve n yerine evren gözlem sayısı N kullanılır.

Buna göre evrenin standart sapması formülü;

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Evrenin varyans formülü;

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Merkezi Değişim Ölçüleri

Uyarı: Standart sapma formülünün paydasında yer alan n yerine bazı kaynaklarda $n-1$ kullanılmaktadır. Ancak gözlem sayısı çok olduğunda bunun önemi yoktur.

Merkezi Değişim Ölçüleri

Aşağıda altı işmakinesinin saatlik hafriyat miktarları (ton) verilmiştir. Verilen dizisinin standart sapmasını hesaplayalım [3].

Örnek:

X_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
7	-2,5	6,25
7	-2,5	6,25
9	-0,5	0,25
10	0,5	0,25
11	1,5	2,25
13	3,5	12,25
57		27,50

Merkezi Değişim Ölçüleri

Örnek:

- Öncelikle aritmetik ortalama hesaplanır;

$$\overline{X} = \frac{\sum X_i}{n} = \frac{57}{6} = 9,5 \text{ ton}$$

- Sonra $\sum_{i=1}^{n} (X_i - \overline{X})^2 = 27,50$ bulunur,
- Formül uygulanarak, standart sapma hesaplanır.

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{27,50}{6}} = 2,14 \text{ ton}$$

Merkezi Değişim Ölçüleri

Diziyi oluşturan gözlem değerlerinin, aritmetik ortalama $9,5$ ton'dan ortalama olarak 2,14 ton'luk sapma gösterdikleri söylenebilir.

Hesaplanan standart sapma küçüldükçe değişkenliğin azaldığı, dolayısıyla gözlem değerlerini birbirine daha çok yaklaştığısöylenebilir. Örneğin, bir serinin $\bar{X} = 75$ ve $s = 0$ olsun bu serinin hiç değişkenliği yok demektir. Ancak bu, gözlem değerlerinin birbirine eşit olması durumunda görülür. Oysa gerçek yaşamda bu tür serilerle pek karşılaşılmaz. Yine de standart sapmanın sıfıra yakın değer alması değişkenliğin az olduğunu göstermektedir.

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

Aşağıdaki basit seride varyans ve standart sapmayı hesaplayınız.

$$4, 5, 8, 15, 18, 22$$

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

Bu veri kümesini kitle gibi düşünürsek kullanmamız gereken formül şudur:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

Öncelikle kitle ortalamasını yani μ'yü hesaplayalım.

Basitçe tüm gözlemleri toplayarak gözlem sayısına bölüyoruz.

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

$\mu = \frac{4+5+8+15+18+22}{6} = 12$

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

Sonra formüle bakıyoruz, dediğimiz gibi formüller konuşurlar ve bize sırasıyla ne yapmamız gerektiğini söylerler.

Şimdi kitle ortalaması μ'yü 12 olarak hesapladık.

Her bir gözlem değerini tek tek bu ortalama değerden saptıracağız.

Bunun için bir tablo oluşturalım:

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

$X_i - \mu$
-8
-7
-4
3
6
10

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

Sonra bu değerlerin tek tek karelerini alıp toplayacağız ve bulduğumuz toplamı n'e yani 6 'ya bölerek varyansı bulmuş olacağız.

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

$(X_i - \mu)^2$
64
49
16
9
36
100

Merkezi Değişim Ölçüleri

Sıra sizde Örnek:

$$\sigma^2 = \frac{64 + 49 + 16 + 9 + 36 + 100}{6} = \frac{274}{6} = 45.667$$

Standart sapma ise varyansın karekökü olduğuna göre;

$$\sigma = 6.758$$

'dir.

Merkezi Değişim Ölçüleri

Frekans ve gruplandırılmış serilerde standart sapma:

- Öncelikle serini aritmetik ortalaması ($\overline{X}$) hesaplanır.
- Hesaplanan aritmetik ortalama, her bir gözlem değerinden çıkarılır ve bu farkların toplamı alınır. $\sum_{i=1}^{n}(X_i - \overline{X})^2 n_i$
- İzleyen formüller uygulanarak standart sapma ve varyans hesaplanır.

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2 . n_i}{\sum n_i}} \text{ varyans ise; } s^2 = \frac{\sum \left[(X_i - \overline{X})^2 . n_i \right]}{\sum n_i} \text{ dır.}$$

Bu formüllerdeki standart sapma ve varyans örnek içindir. Evren standart sapma ve varyans formülleri yazılırken $\overline{X}$ yerine evren aritmetik ortalaması μ ve n yerine evren gözlem sayısı N konur.

Merkezi Değişim Ölçüleri

Varyansın Formülünü İnceleyelim

Varyans kelimesi “variation” yani değişim anlamına gelmektedir.

İstatistiksel olarak yorumlarsak, bir veri kümesindeki gözlemlerin ortaladan uzaklıklarının bir ölçüsüdür diyebiliriz.

Varyansı büyük bir seride birimlerin ortalamadan farklı (ortalamaya uzak) değerlere sahip olduğunu görürüz.

Varyansın küçük olması ise gözlemlerin ortalama dolayında dağılım gösterdiği durumlarda meydana gelir.

Merkezi Değişim Ölçüleri

Bu anlatılanlardan şunu anlamalıyız, eğer bir serideki gözlem birimleri ortalamadan uzakta çok farklı değerler alıyorlarsa bu serinin varyansı (değişimi, yayılımı), ortalama dolayında bir dağılıma sahip olan bir veri kümesinin varyansına göre daha büyüktür.

Formülün paydasında “n” vardır, yani birim sayısı arttıkça varyans küçülmektedir, varyansın yani değişimin küçülmesi ise ileriki konularda anlatacağımız parametre tahminlerinin daha isabetli daha az hata ile yapılmasını sağlayacak bir unsurdur.

Merkezi Değişim Ölçüleri

Değişim Katsayısı:

Bir seriyi özetlemede sadece ortalamanın yeterli olamayacağından bahsetmiştik. Değişim katsayısı, ortalama ile varyansı birlikte kullanarak serilerin değişimlerini ölçmeyi ve farklı türden olsalar bile kıyaslama imkânı sağlayan bir ölçüttür. Değişim katsayısının formülü aşağıdaki gibidir.

$$DK = \frac{\sigma}{\mu} \times 100$$

ya da

$$DK = \frac{s}{\bar{X}} \times 100$$

Merkezi Değişim Ölçüleri

Değişim katsayısı ölçü birimlerinden bağımsızdır.

Oransal yapısından dolayı, pay ve paydasında aynı cins ve büyüklükten değerler birbirini götüreceği için basit seri, sıklık serisi ve sınıflanmış seriler arasındaki cins ve büyüklük farklılığını ortadan kalkar.

Merkezi Değişim Ölçüleri

DK’sı küçük olan serilerin, diğerlerine göre daha az değişken olduğu (daha homojen yani ortalamaya yakın dağılıma sahip) birimlerden oluştuğu söylenir.

Çünkü standart sapma küçüldükçe *DK* da küçülür. Standart sapmanın küçük olması ise ancak ortalama etrafındaki saçılımın $\bar{X}$ ’ya doğru çekilmesiyle mümkündür birimler $\bar{X}$’dan uzaklaştıkça standart sapma da büyümektedir.

Merkezi Değişim Ölçüleri

Örnek:

İki hisse senedi olsun. Bu hisselerin 1 aylık ortalama getirileri ve standart sapmaları aşağıdaki gibi ise, siz bir portföy yöneticisi olarak riski seven bir yatırımcıya hangi hisse senedini tavsiye edersiniz?

$$\mu_1 = 70, \quad \sigma_1 = 6, \quad \mu_2 = 25, \quad \sigma_2 = 3$$

Merkezi Değişim Ölçüleri

Örnek:

$$DK_1 = \frac{6}{70} \times 100 = 8,57$$

$$DK_2 = \frac{3}{25} \times 100 = 12$$

Merkezi Değişim Ölçüleri

Chebyshev Teoremi

Ortalama ve standart sapmanın birlikte kullanıldığı bir durum da, verilerin yüzde kaçının hangi aralıkta bulunduğunu ölçmeye yarayan Chebyshev Teoremi'dir.

Ortalaması μ, standart sapması σ olan bir veri kümesindeki gözlemlerin $\mu \pm k\sigma$ aralığına düşenlerin oranı en az $1 - \frac{1}{k^2}$'dir. Burada k, 1'den büyük bir sayıdır.

Ödev

1- 12, 15, 4, 8, 4 sayılarından oluşan serinin varyans ve standart sapma değerlerini hesaplayınız?

2- Risk almak istemeyen bir yatırımcı için hangi hisse senedini önerirsiniz?

   A hisse senedi ortalama getiri 14 , standart sapması 9.

   B hisse senedi ortalama getiri 12 standart sapması 4.

5. Hafta: Olasılık

İpucu : Zar , Madeni Para

Kaynaklar I

[1] K. Mert Çubukçu,
"Planlamada ve Coğrafyada Temel İstatistik ve Mekansal İstatistik",
Ankara: Nobel Akademik Yayıncılık Eği. Dan. Tic. Ltd. Şti., (2015).

[2] A. Özmen, F. Er, M. Atlas, E. Şıklar,
"İstatistik (AÖF)",
Eskişehir: Anadolu Üniversitesi Yayınları, (2012).

[3] L. İşbilen Yücel,
"İstatistik Maliye Uzaktan Eğitim",
İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi Ders Notu.

[4] Ö. Serper,
"Uygulamalı İstatistik 1",
Bursa: Ezgi Kitapevi, (2004).

Kaynaklar II

[5] Murat Komisyon,
"İstatistik",
Murat Açıköğretim Yayınları, (2004).

[6] N. Gürsakal, A. Oğuzlar,
"Betimsel İstatistik",
Dora Yayıncılık, (2019).

[7] Y. Baykul, C. O. Güzeller,
"Sosyal Bilimler için İstatistik Uygulamaları",
Ankara: Pegem Akademi, (2014).