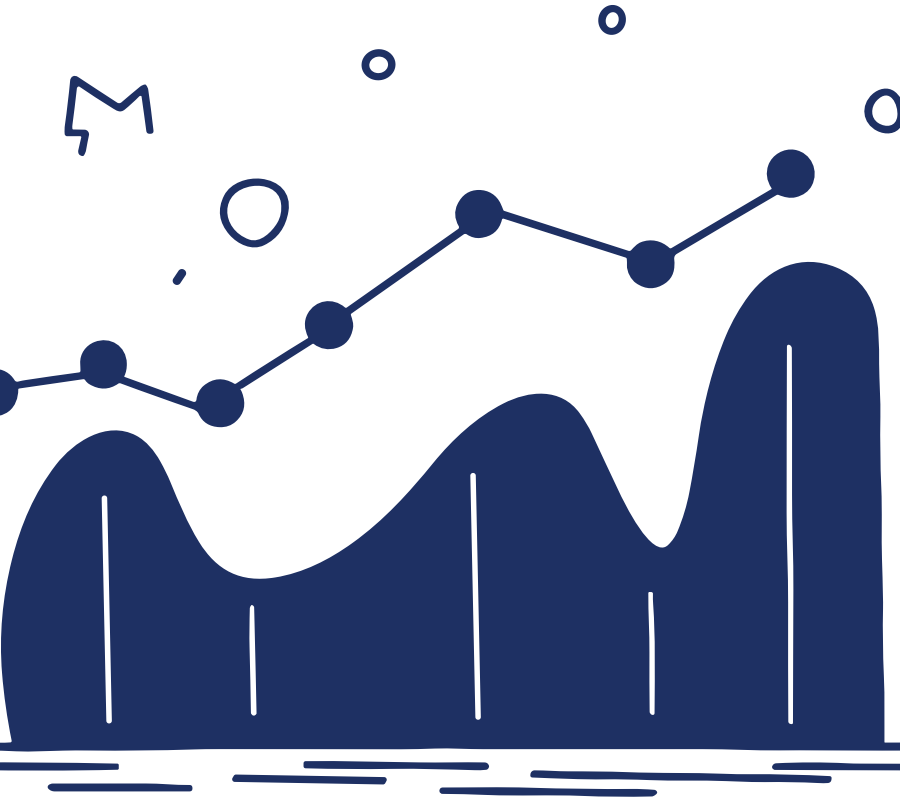


Ortalama eřitleri ve Merkezi Dağılım Ölüleri

Do. Dr. Ömer Faruk Renber





Bölüm 1

Merkezi Eğilim Ölçülerine Giriş

Merkezi eğilim ölçüleri, bir veri setinin genel karakteristiğini tek bir sayı ile özetlemeye yarayan istatistiksel araçlardır. Bu ölçüler, verilerin hangi değer etrafında toplandığını gösterir ve veri setinin genel yapısını anlamamıza yardımcı olur.

Merkezi Eğilim Nedir?

Merkezi eğilim, bir veri setindeki gözlemlerin hangi değer etrafında yoğunlaştığını gösteren istatistiksel bir kavramdır. Bu kavram, verilerin "tipik" veya "ortalama" değerini temsil eder ve veri setinin genel eğilimini anlamak için kullanılır.

Merkezi eğilim ölçüleri, büyük veri setlerini özetlemek ve farklı veri gruplarını karşılaştırmak için vazgeçilmezdir. Araştırmacılar ve analistler, bu ölçüleri kullanarak karmaşık veri yapılarını basit ve anlaşılır hale getirebilirler.

Tanım

Verilerin "ortalama" veya "tipik" değerini gösterir

Özellik

Verilerin merkezinde toplanan bir değer olarak tanımlanır



En Yaygın Ortalama Türleri

$$\frac{f}{dx}$$

Aritmetik Ortalama

Tüm değerlerin toplamının, gözlem sayısına bölünmesiyle elde edilen en yaygın kullanılan ortalama türüdür. Nicel verilerde tercih edilir.



Medyan (Ortanca)

Verilerin sıralanmış halinde tam ortasında bulunan değerdir. Uç değerlerden etkilenmediği için çarpık dağılımlarda önerilir.



Mod (Tepe Değer)

Veri setinde en sık tekrar eden değerdir. Hem nicel hem de nitel verilerde kullanılabilir.

Aritmetik Ortalama Nedir?

Aritmetik ortalama, bir veri setindeki tüm değerlerin toplamının, toplam gözlem sayısına bölünmesiyle hesaplanan en temel ve en yaygın kullanılan merkezi eğilim ölçüsüdür. Sembolik olarak $\bar{x}$ veya μ ile gösterilir.

Formül

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

Burada n gözlem sayısını, x_i ise her bir gözlem değerini temsil eder.

Örnek Hesaplama

Veri Seti: {5, 10, 15}

Hesaplama: $(5 + 10 + 15) / 3 = 30 / 3 = 10$

Sonuç: Aritmetik Ortalama = **10**

Aritmetik ortalama, veri setindeki tüm değerleri dikkate aldığı için matematiksel işlemlerde kullanışlıdır, ancak uç değerlerden etkilenebilir.

Medyan (Ortanca) Nedir?

Medyan, bir veri setinin küçükten büyüğe sıralandığında tam ortasında yer alan değerdir. Veri sayısı tek ise ortadaki değer, çift ise ortadaki iki değer aritmetik ortalaması medyana verir.

01

Verileri Sırala

Veri setini küçükten büyüğe doğru sırala

02

Ortayı Bul

Gözlem sayısı tek mi çift mi kontrol et

03

Medyanı Belirle

Tek ise ortadaki değer, çift ise ortadaki iki değer ortalamasını al

Örnek 1 (Tek Sayı)

Veri: {5, 10, 15}

Medyan: 10

Örnek 2 (Çift Sayı)

Veri: {5, 10, 15, 20}

Medyan: $(10+15)/2 = 12.5$

Önemli Not: Medyan, aritmetik ortalamanın aksine uç değerlerden (outlier) etkilenmez, bu nedenle çarpık dağılımlarda daha güvenilir bir merkezi eğilim ölçüsüdür.



Mod (Tepe Değer) Nedir?

Mod, bir veri setinde en sık tekrar eden değerdir. Bir veri setinde birden fazla mod olabilir (bimodal, multimodal) veya hiç mod olmayabilir. Mod, özellikle kategorik verilerin analizinde ve frekans dağılımlarında önemli bir ölçüdür.

Mod Türleri

- **Unimodal:** Tek bir mod değeri vardır
- **Bimodal:** İki farklı mod değeri vardır
- **Multimodal:** İki'den fazla mod değeri vardır
- **Amodal:** Hiçbir değer tekrar etmez

Örnek

Veri Seti: {5, 10, 10, 15}

Mod: 10

Çünkü 10 değeri veri setinde iki kez tekrar etmektedir ve en yüksek frekansa sahiptir.

Mod, nitel (kategorik) verilerde de kullanılabilen tek merkezi eğilim ölçüsüdür. Örneğin, en popüler renk, en çok tercih edilen marka gibi.

Ortalama, Medyan ve Mod'un Dağılımdaki Konumları

Simetrik Dağılım

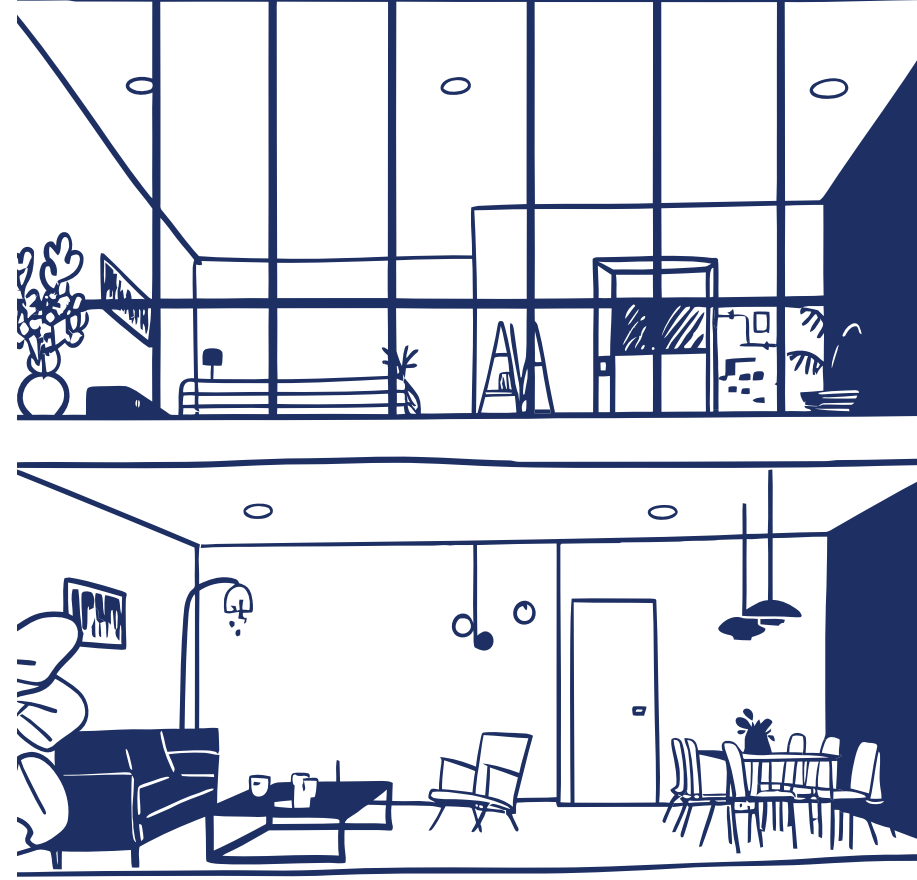
Normal dağılımda ortalama, medyan ve mod aynı noktadadır. Bu, verilerin mükemmel bir simetri gösterdiğini ifade eder.

Sağa Çarpık Dağılım

$\text{Mod} < \text{Medyan} < \text{Ortalama}$. Yüksek değerler ortalamanın sağa kaymasına neden olur.

Sola Çarpık Dağılım

$\text{Ortalama} < \text{Medyan} < \text{Mod}$. Düşük değerler ortalamanın sola kaymasına neden olur.

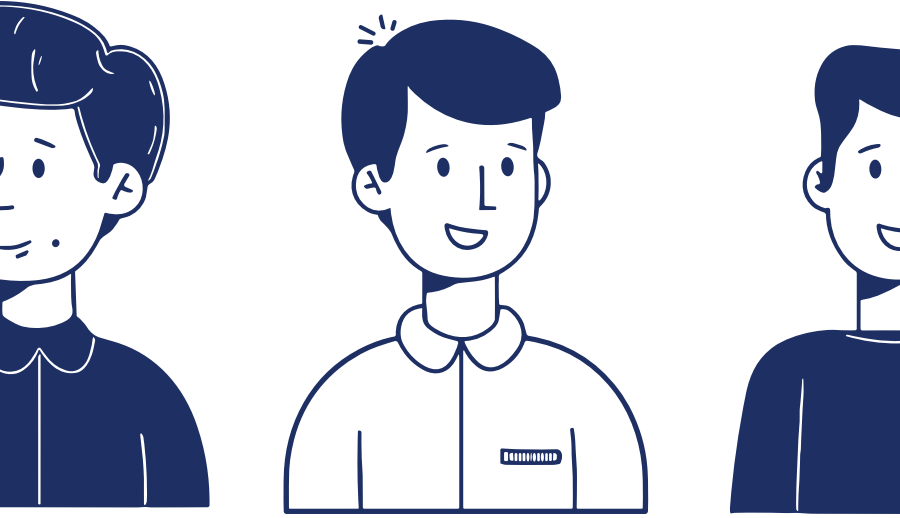




Bölüm 2

Merkezi Dağılım Ölçüleri

Merkezi eğilim ölçüleri veri setinin merkezini gösterirken, merkezi dağılım ölçüleri verilerin bu merkez etrafında nasıl yayıldığını, ne kadar değişkenlik gösterdiğini ortaya koyar.



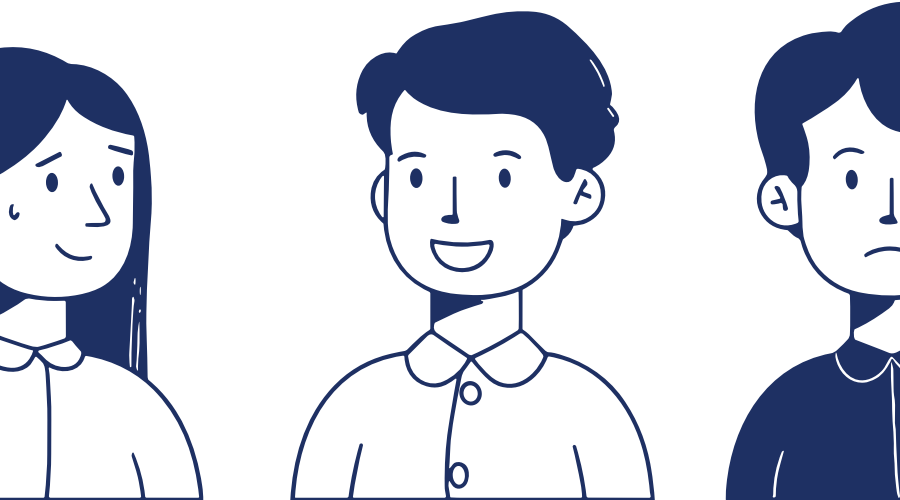
Merkezi Dağılım Ölçüleri Nedir?

Merkezi dağılım ölçüleri, bir veri setindeki gözlemlerin merkezi eğilim ölçüsünden ne kadar uzaklaştığını, yani verilerin ne kadar yayıldığını veya dağıldığını gösteren istatistiksel araçlardır.

Kritik Önemi: Ortalama tek başına veri setinin yapısını tam olarak yansıtmaz. Verilerin ne kadar homojen veya heterojen olduğunu anlamak için dağılım ölçüleri şarttır.

İki veri seti aynı ortalamaya sahip olabilir ancak verilerin dağılımı çok farklı olabilir. Bu nedenle, sadece merkezi eğilim ölçüleri veri setini tam olarak tanımlamak için yeterli değildir.

"İstatistikte ortalama, hikayenin sadece başlangıcıdır. Dağılım ölçüleri, hikayenin geri kalanını anlatır."



Temel Dağılım Ölçüleri



Ortalama Sapma

Her bir gözlemin ortalamadan mutlak sapmasının ortalamasıdır. Basit ve sezgisel bir dağılım ölçüsü olarak verilerin ortalamadan ortalama uzaklığını gösterir.



Varyans

Her bir gözlemin ortalamadan sapmasının karesinin ortalamasıdır. En yaygın kullanılan dağılım ölçüsü olup istatistiksel analizlerde temel oluşturur.



Standart Sapma

Varyansın karekökü olarak hesaplanır. Orijinal veri birimiyle aynı birimde olduğu için yorumlanması ve anlaşılması daha kolaydır.



Değişim Katsayısı

Standart sapmanın ortalamaya oranının yüzde olarak ifadesidir. Farklı birimlere veya ölçeklere sahip veri setlerinin değişkenliğini karşılaştırmak için kullanılır.

Ortalama Sapma Nedir?

Ortalama sapma (mean deviation), veri集中的 her bir gözlemin aritmetik ortalamadan ne kadar uzaklaştığının mutlak değeriinin ortalamasıdır. Bu ölçü, verilerin ortalama etrafındaki yayılımını basit bir şekilde ifade eder.

Formül

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

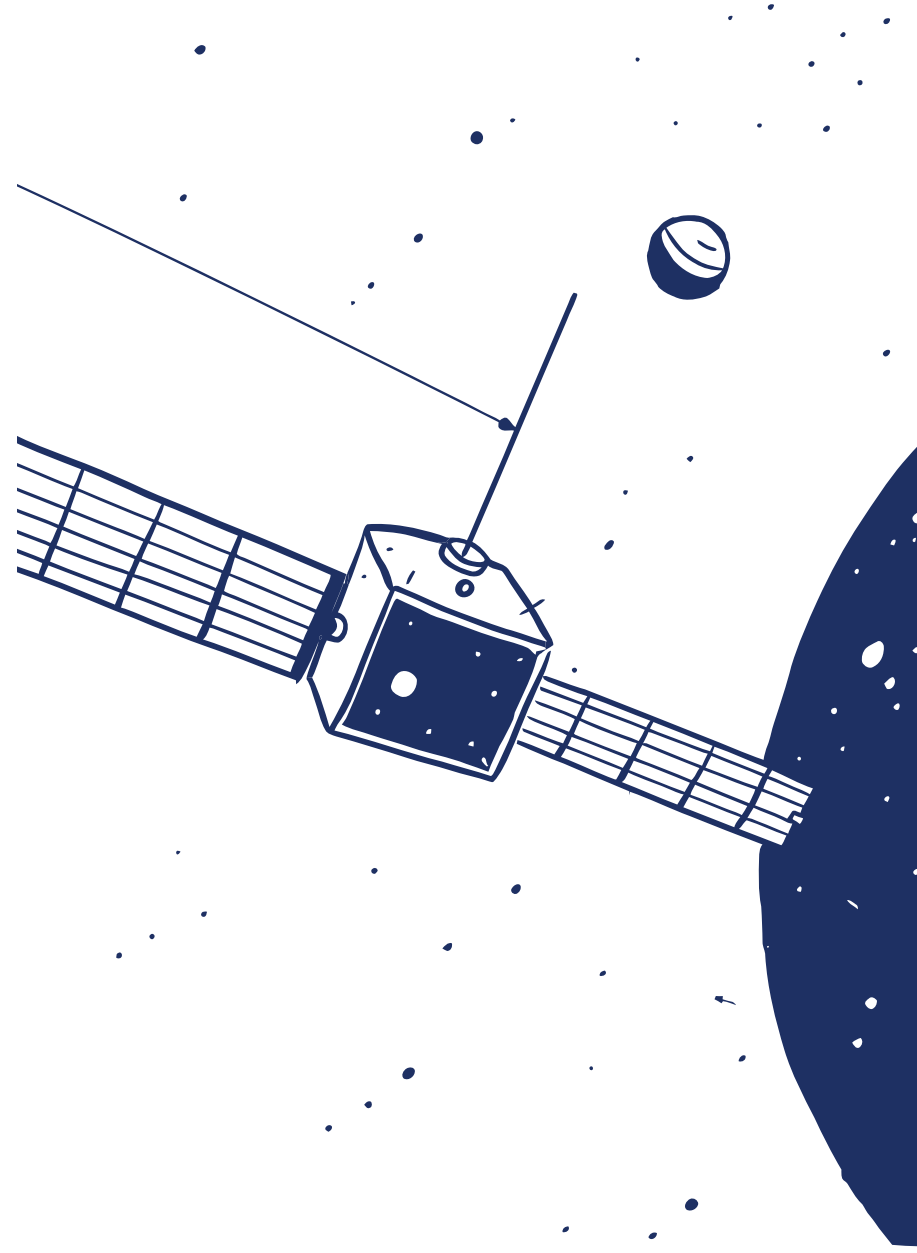
Ortalama sapma hesaplanırken sapmaların işaretleri önemli değildir, mutlak değeri alınır. Bu sayede pozitif ve negatif sapmalar birbirini götürmez.

Avantajları

- Basit ve anlaşılır bir ölçüdür
- Hesaplaması kolaydır
- Tüm gözlemleri dikkate alır

Dezavantajları

- Mutlak değeri kullanımı matematiksel işlemleri zorlaştırır
- İleri istatistiksel analizlerde kullanımı sınırlıdır
- Varyans ve standart sapma kadar yaygın değildir



Varyans Nedir?

Varyans, bir veri setindeki her bir gözlemin aritmetik ortalamadan farkının karesinin ortalamasıdır. Sembolik olarak s^2 (örneklem) veya σ^2 (kitle) ile gösterilir.

Kitle Varyansı

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

Tüm kitleyi temsil eden parametredir. Payda olarak N kullanılır.

Örneklem Varyansı

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Örneklemden hesaplanan istatistiktir. Payda olarak $n - 1$ kullanılır (Bessel düzeltmesi).

Varyansın Önemi

Varyans, verilerin ortalama etrafındaki yayılımını sayısal olarak ifade eder. Yüksek varyans, verilerin ortalamadan uzak olduğunu; düşük varyans ise verilerin ortalamaya yakın toplandığını gösterir.

Neden Kare Alınır?

Sapmaların karesi alınarak negatif ve pozitif sapmalar birbirini götürmez. Ayrıca kare alma işlemi, büyük sapmaları daha fazla vurgular ve matematiksel işlemleri kolaylaştırır.



Standart Sapma Nedir?

Standart sapma, varyansın karekökü alınarak elde edilen ve veri setindeki gözlemlerin ortalamadan ortalama ne kadar saptığını gösteren en yaygın kullanılan dağılım ölçüsüdür. Sembolik olarak s (örneklem) veya σ (kitle) ile gösterilir.

Formüller

Kitle Standart Sapması:

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

Örneklem Standart Sapması:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Yorumlama

Standart sapma, orijinal verilerle aynı birimde olduğu için yorumlanması varyansa göre çok daha kolaydır.

Normal dağılımda:

- **%68** veri: $\mu \pm \sigma$ aralığında
- **%95** veri: $\mu \pm 2\sigma$ aralığında
- **%99.7** veri: $\mu \pm 3\sigma$ aralığında

Önemli: Düşük standart sapma homojen bir veri setini, yüksek standart sapma ise heterojen bir veri setini işaret eder.

Varyans ve Standart Sapma Arasındaki Fark

Varyans ve standart sapma, her ikisi de verilerin dağılımını ölçen araçlar olmakla birlikte, aralarında önemli farklar bulunmaktadır. Bu farkları anlamak, doğru ölçüyü seçmek açısından kritiktir.

Varyans (σ^2 veya s^2)

- Sapmaların **karelerinin** ortalaması
- Birimi: **Orijinal veri biriminin karesi** (örn: cm^2)
- Matematiksel işlemlerde tercih edilir
- Değeri standart sapmadan büyüktür
- Yorumlanması daha zordur

Standart Sapma (σ veya s)

- Varyansın **karekökü**
- Birimi: **Orijinal veri birimiyle aynı** (örn: cm)
- Pratik uygulamalarda tercih edilir
- Değeri varyansın kareköküdür
- Yorumlanması daha kolay ve sezgiseldir

Örnek

Boy ölçümlerinde varyans 25 cm^2 ise, standart sapma 5 cm'dir. Standart sapmanın orijinal birimde (cm) olması yorumlamayı kolaylaştırır.

Sonuç: Her iki ölçü de aynı bilgiyi farklı şekillerde ifade eder. Standart sapma günlük analizlerde, varyans ise teorik ve ileri istatistiksel hesaplamalarda daha yaygındır.

Değişim Katsayısı Nedir?

Değişim katsayısı (coefficient of variation - CV), standart sapmanın aritmetik ortalamaya oranının yüzde olarak ifadesidir. Bu ölçü, farklı birimlere veya ölçeklere sahip veri setlerinin göreceli değişkenliğini karşılaştırmayı mümkün kılar.

Formül

$$CV = \frac{s}{\bar{x}} \times 100\%$$



Birimsiz Ölçü

Yüzde olarak ifade edildiği için birimsizdir ve farklı ölçekli verileri karşılaştırmaya olanak tanır.



Göreceli Değişkenlik

Mutlak değil, göreceli (relatif) değişkenliği ölçer, bu da farklı ortalamalara sahip veri setlerinin karşılaştırılmasını sağlar.



Kullanım Alanları

Kalite kontrol, finansal analizler, bilimsel araştırmalarda yaygın olarak kullanılır.

Örnek 1: Homojen Veri

Ortalama = 100, Standart Sapma = 5

$$CV = (5/100) \times 100 = \%5$$

Düşük CV, düşük değişkenlik gösterir.

Örnek 2: Heterojen Veri

Ortalama = 100, Standart Sapma = 30

$$CV = (30/100) \times 100 = \%30$$

Yüksek CV, yüksek değişkenlik gösterir.

Merkezi Eğilim ile Merkezi Dağılım Arasındaki Fark

Merkezi eğilim ve merkezi dağılım ölçüleri, bir veri setini tanımlamak için birlikte kullanılan ama farklı amaçlara hizmet eden kavramlardır. Bu iki kavram arasındaki farkı anlamak, istatistiksel analizde kritik öneme sahiptir.

Merkezi Eğilim Ölçüleri

- **Amaç:** Verilerin "merkez" değerini belirlemek
- **Soru:** "Veriler hangi değer etrafında toplanıyor?"
- **Ölçüler:** Ortalama, Medyan, Mod
- **Bilgi:** Veri setinin tipik veya ortalama değeri
- **Tek başına:** Verilerin yayılımı hakkında bilgi vermez

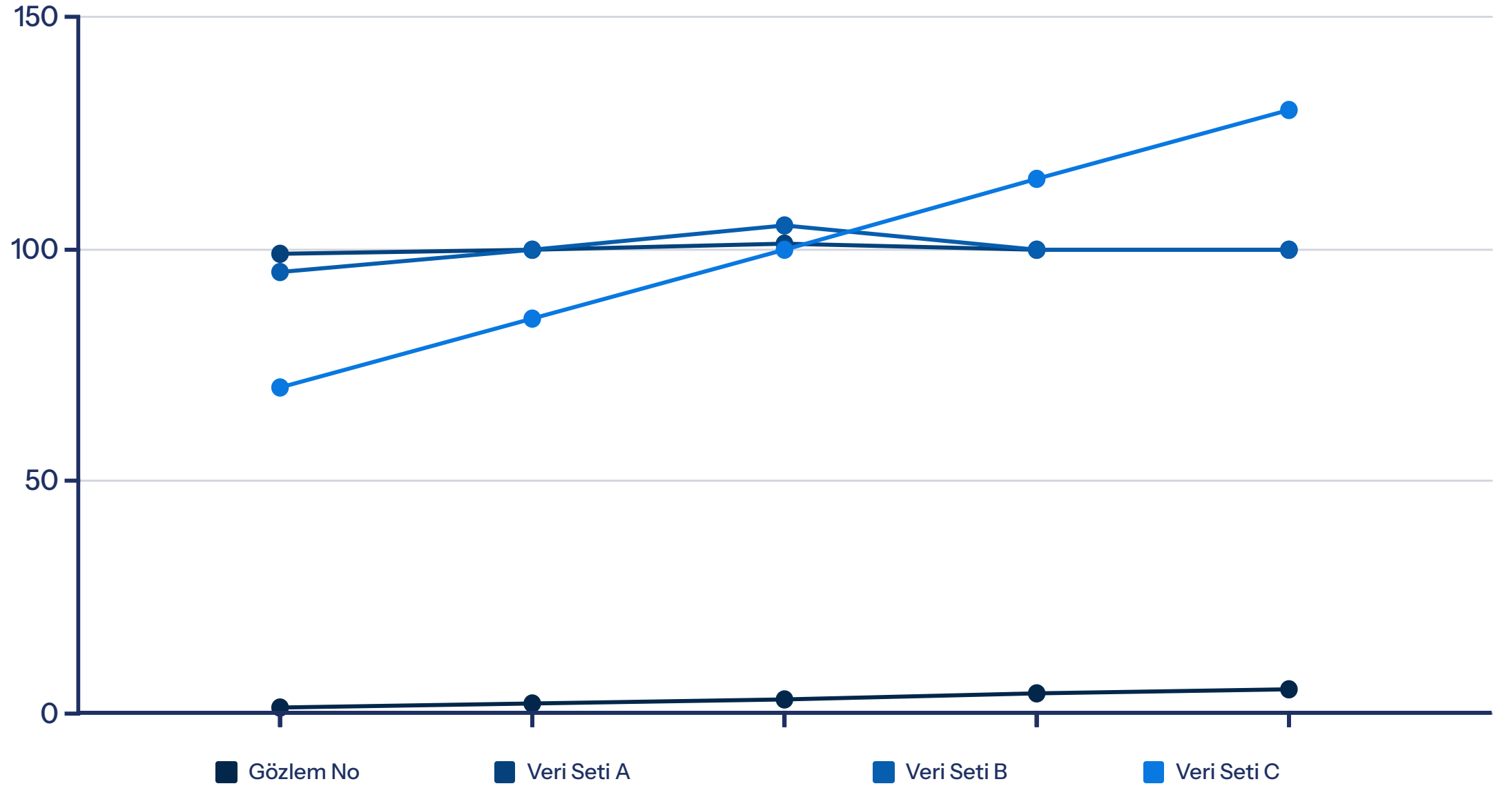
Merkezi Dağılım Ölçüleri

- **Amaç:** Verilerin ne kadar yayıldığını ölçmek
- **Soru:** "Veriler ne kadar değişkenlik gösteriyor?"
- **Ölçüler:** Ortalama Sapma, Varyans, Standart Sapma, Değişim Katsayısı
- **Bilgi:** Veri setinin homojenliği veya heterojenliği
- **Tek başına:** Verilerin merkez değeri hakkında bilgi vermez

Önemli: İkisi birlikte kullanıldığında veri setinin tam ve eksiksiz bir resmini ortaya koyar. Merkezi eğilim "nerede", merkezi dağılım ise "ne kadar yayılmış" sorularını cevaplar.

Örnek: Aynı Ortalama, Farklı Dağılım

Aşağıdaki örnek, neden merkezi eğilim ölçülerinin tek başına yeterli olmadığını açıkça göstermektedir. Üç farklı veri seti aynı ortalamaya (100) sahiptir, ancak dağılımları tamamen farklıdır.



100

Veri Seti A

Standart Sapma: ≈ 0.7

Çok dar dağılım, yüksek homojenlik

100

Veri Seti B

Standart Sapma: ≈ 4.2

Orta düzeyde dağılım

100

Veri Seti C

Standart Sapma: ≈ 21.2

Çok geniş dağılım, yüksek heterojenlik

Bu örnek, aynı ortalamaya sahip veri setlerinin tamamen farklı karakteristikler gösterebileceğini kanıtlamaktadır. Dağılım ölçüleri olmadan veri setinin gerçek yapısını anlamamız imkansızdır.

Üç Farklı Dağılım: Aynı Ortalama, Farklı Standart Sapma

Yukarıdaki görselleştirme, üç farklı normal dağılımı göstermektedir. Her üç dağılımın da ortalaması aynı noktadadır (merkez çizgi), ancak standart sapmaları farklıdır.

Dar Dağılım ($\sigma = \text{küçük}$)
Veriler ortalamaya çok yakın, yüksek konsantrasyon, düşük değişkenlik, yüksek tutarlılık



Orta Dağılım ($\sigma = \text{orta}$)

Veriler ortalama etrafında dengeli yayılım, orta düzeyde değişkenlik, tipik dağılım paterni

Geniş Dağılım ($\sigma = \text{büyük}$)

Veriler geniş alana yayılmış, yüksek değişkenlik, düşük öngörülebilirlik, heterojen yapı

Bu görsel karşılaştırma, istatistiksel analizde hem merkezi eğilim hem de dağılım ölçülerinin birlikte değerlendirilmesinin neden hayati önem taşıdığını açıkça ortaya koymaktadır.

Sonuç: Ortalama ve Dağılım Ölçüleri Birlikte Anlamlıdır

Temel İlkeler

Bir veri setini tam olarak anlamak ve yorumlamak için hem merkezi eğilim hem de merkezi dağılım ölçülerinin birlikte kullanılması gerekmektedir. Bu bütünsel yaklaşım, istatistiksel analizin temelidir.

Ortalama Tek Başına Yanıltıcıdır

Sadece ortalamaya bakarak veri seti hakkında sağlıklı yorum yapılamaz. İki veri seti aynı ortalamaya sahip olabilir ama tamamen farklı yapılara sahip olabilir.

Dağılım Çeşitliliği Gösterir

Dağılım ölçüleri, verilerin homojen mi yoksa heterojen mi olduğunu, değişkenliğin derecesini ve güvenilirliği ortaya koyar.

Birlikte Eksiksiz Resim

Merkezi eğilim "nerede" sorusunu, dağılım ise "nasıl" sorusunu cevaplar. İkisi birlikte veri setinin eksiksiz karakterizasyonunu sağlar.

Pratik Öneriler

İstatistiksel Analizde Altın Kural

Her veri analizi raporunda şunlar mutlaka yer almalıdır:

- En az bir merkezi eğilim ölçüsü (tercihen ortalama ve medyan)
- En az bir dağılım ölçüsü (tercihen standart sapma)
- Uygun görsel sunumlar (histogram, kutu grafiği vb.)

Kritik Uyarılar

- Uç değerler (outliers) ortalamayı etkileyebilir
- Çarpık dağılımlarda medyan daha güvenilirdir
- Standart sapma sıfıra yaklaşırsa veriler çok homogendir
- Değişim katsayısı %30'u aşarsa yüksek değişkenlik vardır

Sonuç: İstatistiksel düşünme, verileri sadece rakamlar olarak değil, anlamlı bilgi kaynakları olarak görmeyi gerektirir. Bu perspektif, doğru kararlar almak için vazgeçilmezdir.