

# Merkezi Limit Teoremi(CLT)

Bir veri setinin (popülasyonun) orijinal dağılımı ne kadar garip veya düzensiz olursa olsun, o veri setinden alacağın yeterince büyük ve rastgele örneklemelerin ortalaması, her zaman "Normal Dağılım"a (Çan Eğrisi) uyar.

## 1. Temel Mantık: Kaostan Düzen Doğar

Normalde her veri seti düzgün (normal) dağılmaz. Örneğin, bir şirketteki maaşlar veya insanların günlük adım sayıları genellikle çarpık dağılır (sağa veya sola yatık).

Ancak CLT der ki:

- O veri setinden rastgele küçük gruplar (örneklem) seç.
- Bu grupların ortalamasını al.
- Bu işlemi binlerce kez tekrarla.
- **Sonuç:** Elde ettiğin bu binlerce ortalamayı grafiğe dökersen, **kusursuz bir çan eğrisi (Normal Dağılım)** görürsün.

## 2. Zar Örneği ile Somutlaştıralım

Bu durumu en iyi **zar atma** örneği açıklar:

- **Tek bir zar atarsan:** 1, 2, 3, 4, 5 veya 6 gelme olasılığı eşittir (Düz/Uniform dağılım). Bir tepe noktası yoktur.
- **İki zar atıp ortalamasını alırsan:** Uç değerler (1 ve 6) azalmaya, ortalamalar (3.5 civarı) artmaya başlar.
- **30 zar atıp ortalamasını alırsan:** Ortalamaların neredeyse tamamı 3.5 civarında toplanır ve grafik **Normal Dağılım** şeklini alır.

## 3. "30" Kuralı (Sihirli Sayı)

İstatistikçiler genellikle  $n \geq 30$  kuralını kabul ederler. Yani, çektiğin örneklem sayısı (n) 30 veya daha fazlaysa, orijinal veri dağılımı ne olursa olsun, örneklem ortalamalarının dağılımını normal kabul edebiliriz.

Python



```
import numpy as np
import matplotlib.pyplot as plt

# 1. ADIM: Kaotik (Düzgün) Bir Popülasyon Oluşturma
# 1 ile 6 arasında rastgele 100.000 adet sayı (Zar atma simülasyonu)
# Bu dağılım 'Normal' değildir, dümdüzdür (Uniform).
populasyon = np.random.randint(1, 7, 100000)

# 2. ADIM: Örneklem Ortalamalarını Toplama
orneklem_ortalamlari = []
simulasyon_sayisi = 1000 # Bu işlemi 1000 kere tekrarla
n = 30 # Her seferinde 30 tane zar seç (n=30 kuralı)

for i in range(simulasyon_sayisi):
    # Popülasyondan rastgele n tane sayı seç
    orneklem = np.random.choice(populasyon, n)
    # Bu grubun ortalamasını al ve listeye ekle
    orneklem_ortalamlari.append(np.mean(orneklem))

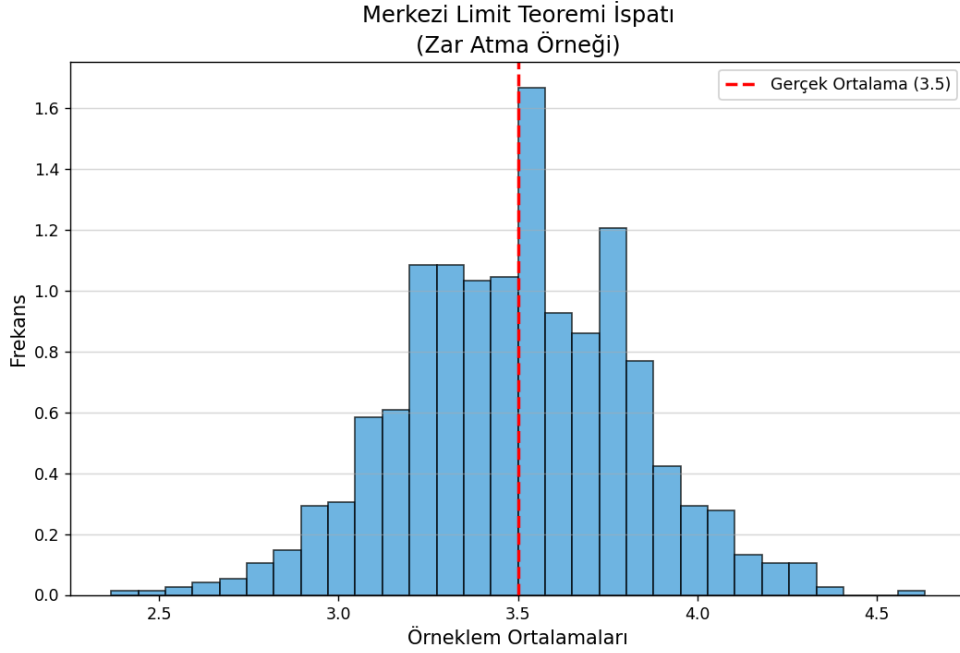
# 3. ADIM: Görselleştirme
plt.figure(figsize=(10, 6))

# Histogramı çizelim
plt.hist(orneklem_ortalamlari, bins=30, density=True, alpha=0.7, color='#3498db',

# Teorik ortalamayı (Zar için 3.5) kırmızı çizgiyle gösterelim
plt.axvline(3.5, color='red', linestyle='dashed', linewidth=2, label='Gerçek Ortala

plt.title('Merkezi Limit Teoremi İspatı\n(Zar Atma Örneği)', fontsize=14)
plt.xlabel('Örneklem Ortalamaları', fontsize=12)
plt.ylabel('Frekans', fontsize=12)
plt.legend()
plt.grid(axis='y', alpha=0.5)

plt.show()
```



### 1. Tepe Noktası Gerçeği Söyler (Tahmin Gücü)

Grafiğin en yüksek noktası (tepe) tam olarak **3.5** üzerindedir.

- **Çıkarım:** "Elimdeki popülasyonun (tüm verinin) ortalamasını bilmiyor olsaydım bile, sadece bu örneklemle bakarak gerçek ortalamanın 3.5 olduğunu %99 güvenle söyleyebilirdim."
- **Anlamı:** Rastgele örneklem, bizi her zaman gerçeğin merkezine götürür. Tepe noktası, **Popülasyon Ortalaması ( $\mu$ )** için en iyi tahmincidir.

### 2. Genişlik Bize "Risk"i Gösterir (Standart Hata)

Grafiğin tabanı 1 ile 6 arasına yayılmaz; muhtemelen **3.0 ile 4.0** arasına sıkışmıştır.

- **Çıkarım:** "Eğer rastgele 30 kişi (veya zar) seçersem, bunların ortalamasının 3.0'dan az veya 4.0'dan çok çıkması çok zordur."
- **Anlamı:** Grafik ne kadar darsa (sivriyse), tahminlerimiz o kadar kesindir. Buna istatistikte **Standart Hata (Standard Error)** denir.

### 3. Uçlar "İmkansız" Gösterir (Aykırı Değer Analizi)

Grafiğin 1.0 veya 6.0 noktalarında hiç çubuk (bar) olmadığını veya çok az olduğunu görürsün.

- **Çıkarım:** "30 tane zar atıp hepsinin 1 gelmesi (ortalama 1) veya hepsinin 6 gelmesi (ortalama 6) teorik olarak mümkün olsa da, pratikte imkansıza yakındır."
- **Anlamı:** Eğer bir gün yaptığın analizde ortalama 1.0 çıkarsa, sistemde bir hile veya hata olduğunu anlarsın. Bu çıkarım, bankalarda **dolandırıcılık tespiti (fraud detection)** için kullanılır.

#### 4. Simetri Güven Verir

Grafik sağa veya sola yatık değil, aynalanmış gibidir.

- **Çıkarım:** "Ortalamadan sapmalar rastgeledir. Pozitif ve negatif yönde hata yapma ihtimalim eşittir."
- **Anlamı:** Verinin davranışı öngörülebilirdir. Bu simetri sayesinde geleceğe yönelik olasılık hesaplarını (Z-tablosu kullanarak) kolayca yapabiliriz.

#### Özetle O Grafik Bize Şunu Söyler:

"Eğer eline bir zar alırsan her şey gelebilir (**Belirsizlik**). Ama eline 30 zar alıp ortalamasına bakarsan, sonuç kesinlikle 3 ile 4 arasında olacaktır (**Belirlilik**). Büyükler ve küçükler birbirini yok etmiştir."

Bu mantık senin "süreç iyileştirme" veya "analiz" projelerinde şu işe yarar: Bir günlük satış rakamın çok düşük (1) veya çok yüksek (6) olabilir. Ama **aylık ortalamana** (30 gün) baktığında o uç değerler törpülenir ve şirketin gerçek performansını (3.5) görürsün.

Bu ortalamayı (ve onun güvenilir olduğunu) bilmek, iş dünyasında ve bilimde bize **"Geleceği Görme"** ve **"Karar Alma"** gücü verir.

Eğer Merkezi Limit Teoremi sayesinde o ortalamaya güvenemeseydik, **stratejik hiçbir karar alamazdık.**

İşte sana bu bilginin sağladığı 3 devasa fayda (Somut İş Örnekleriyle):

#### 1. Tümünü Saymadan Bütünü Yönetirsin (Maliyet ve Zaman Tasarrufu)

Diyelim ki Türkiye genelinde mağazaları olan bir giyim markasının yöneticisisin.

- **Sorun:** "Müşterilerimizin mağazada ortalama ne kadar harcadığını bilmem lazım ki seneye ne kadar stok yapacağımı planlayayım."
- **İmkansız Olan:** 1 milyon müşterinin tek tek fişini incelemek (Aylar sürer).
- **Fayda:** Rastgele 500 fişi seçip ortalamasını alırsın (Örn: 1200 TL). CLT sayesinde bilirsin ki, **tüm Türkiye cironun ortalaması da 1200 TL civarındadır.**
- **Sonuç:** Milyonlarca veriyi işlemeye gerek kalmadan, sadece küçük bir grupla koca şirketin bütçesini **%99 güvenle** planlarsın.

## 2. "Normal" Olanı Belirlersin (KPI ve Performans Belirleme)

Bir şeyin "iyi" mi "kötü" mü olduğunu anlamak için önce "normal"in ne olduğunu bilmen gerekir. O bulduğumuz tepe noktası (ortalama) bizim **referans noktamızdır.**

- **Senaryo:** Bir çağrı merkezinde çalışıyorsun. CLT ile analiz yaptın ve genel çağrı süresi ortalamasının **3.5 dakika** olduğunu buldun.
- **Fayda:** Eğer Ahmet isimli bir çalışan çağrıları ortalama **6 dakikada** bitiriyorsa, onun performansının düşük olduğunu veya eğitime ihtiyacı olduğunu matematiksel olarak kanıtlarsın.
- **Sonuç:** Kişisel görüşlere göre değil, veriye dayalı (Data-Driven) yönetim yaparsın.

## 3. Anormallikleri ve Hileleri Yakalarsın (Risk Yönetimi)

Eğer ortalamanın (tepe noktasının) nerede olması gerektiğini biliyorsan, sapan şeyleri hemen fark edersin.

- **Senaryo:** Bir cips fabrikasında paketlerin 100gr olması gerekiyor. Örneklem aldın ve ortalama **92gr** çıktı.
- **Fayda:** CLT sana der ki: "30 paketin ortalamasının tesadüfen bu kadar düşük çıkması imkansız. Makine bozuk!"
- **Sonuç:** Binlerce hatalı ürün piyasaya sürülmeden üretimi durdurursun. Şirketi devasa bir tazminat riskinden kurtarırsın.