

VERİ BİLİMİ DERSİ

# İstatistiksel Analiz ve A/B Testleri

Hafta 4 · Modül 4

Tanımlayıcı İstatistik, Dağılımlar, Hipotez Testleri ve A/B Testi

**Dr. Murat Altun**

Veri Bilimi ve Yapay Zekâ Eğitimi · 2026

6

Saat

3

Notebook

7+

İstatistik Testi

# İçindekiler

01

## Tanımlayıcı İstatistik

Ortalama · Medyan · Mod · Standart Sapma · Merkezi Eğilim · Yayılım

Slayt 3-6

02

## Dağılımlar ve Korelasyon

Normal Dağılım · Normallik Testleri · Pearson · Spearman

Slayt 5-7

03

## Hipotez Testleri

$H_0/H_1$  · p-value · T-Test · ANOVA · Ki-Kare · Güven Aralıkları

Slayt 8-13

04

## A/B Testi ve Uygulama

Metodoloji · Uygulama · Tip I/II Hata · Notebook'lar · Ödev

Slayt 14-20

$\bar{X}$

Ortalama  
Tüm değerlerin toplamının  
veri sayısına bölümü

$\tilde{M}$

Medyan  
Sıralanmış verinin  
orthadaki değeri

Mo

Mod  
En sık tekrar eden  
değer

$\sigma$

Standart Sapma  
Verinin ortalamadan  
ne kadar saptığı

## Python ile Hesaplama

```
import numpy as np
import pandas as pd

veri = pd.Series([12, 15, 12, 18, 20, 15, 12, 25, 30])
print(f"Ortalama: {veri.mean():.2f}")
print(f"Medyan: {veri.median():.2f}")
print(f"Mod: {veri.mode()[0]}")
print(f"Std: {veri.std():.2f}")
```

## Merkezi Eğilim Ölçüleri

1

### Ortalama (Mean)

Tüm değerlerin aritmetik ortalaması. Aşırı değerlerden etkilenir.

2

### Medyan (Median)

Sıralanmış verinin orta noktası. Aşırı değerlere dayanıklı.

3

### Mod (Mode)

En sık tekrar eden değer. Kategorik veri için idealdir.

4

### Ağırlıklı Ortalama

Her değere farklı ağırlık verilir. Örn: not ortalaması.

## Yayılım Ölçüleri

1

### Varyans ( $\sigma^2$ )

Ortalamadan sapmaların karelerinin ortalaması.

2

### Standart Sapma ( $\sigma$ )

Varyansın karekökü. Orijinal birimle aynı ölçekte.

3

### Aralık (Range)

Max - Min. En basit yayılım ölçüsü, aşırı değerlere hassas.

4

### IQR ( $Q_3 - Q_1$ )

Çeyrekler arası açıklık. Boxplot'un temelini oluşturur.

## Normal Dağılım Nedir?

Doğadaki pek çok olay normal dağılıma uyar: boy, ağırlık, sınav puanları...

- Simetrik, çan şeklinde eğri
- Ortalama = Medyan = Mod
- $\mu$  (ortalama) ve  $\sigma$  (std sapma) ile tanımlanır
- Pek çok istatistiksel testin ön koşulu

## 68-95-99.7 Kuralı

$$\mu \pm 1\sigma$$

%68.3 veri bu aralıkta

$$\mu \pm 2\sigma$$

%95.4 veri bu aralıkta

$$\mu \pm 3\sigma$$

%99.7 veri bu aralıkta

## Çan Eğrisi Özellikleri

- Eğri altındaki toplam alan = 1
- Ortalamadan uzaklaştıkça olasılık azalır
- Z-skoru:  $(X - \mu) / \sigma$  ile standartlaştırılır

- $\sigma$  küçük  $\rightarrow$  sivri eğri (homojen veri)
- $\sigma$  büyük  $\rightarrow$  yayvan eğri (heterojen veri)
- Merkezi limit teoremi:  $n \geq 30 \rightarrow$  örneklem dağılımı  $\approx$  normal

## Neden Normallik Testi?

Pek çok parametrik test (t-test, ANOVA) verinin normal dağıldığını varsayar. Test seçiminden önce normallik kontrolü şart!

## Görsel vs İstatistiksel

- Histogram + KDE → gözle değerlendirme
- Q-Q Plot → doğruya yakınlık
- Shapiro-Wilk →  $p < 0.05$  → normal değil

```
from scipy.stats import shapiro
import matplotlib.pyplot as plt

# Shapiro-Wilk normallik testi
stat, p_value = shapiro(veri)
print(f"Test İstatistik: {stat:.4f}")
print(f"p-value: {p_value:.4f}")

if p_value > 0.05:
    print("✅ Normal dağılım varsayımı reddedilmez")
else:
    print("❌ Normal dağılım değil → non-parametrik test")

# Histogram ile görsel kontrol
plt.hist(veri, bins=20, edgecolor='black', alpha=0.7)
plt.title('Veri Dağılımı')
plt.show()
```

## Pearson vs Spearman Karşılaştırması

Özellik	Pearson (r)	Spearman ( $\rho$ )
Veri Türü	Sürekli, normal dağılım	Sıralı veya sürekli
İlişki Tipi	Doğrusal ilişki ölçer	Monoton ilişki ölçer
Dayanıklılık	Aşırı değerlere hassas	Aşırı değerlere dayanıklı
Kullanım	Yaş-Gelir ilişkisi	Müşteri memnuniyeti sıralaması

## Korelasyon Katsayısı Yorumlama Rehberi

0.00 - 0.19	0.20 - 0.39	0.40 - 0.59	0.60 - 0.79	0.80 - 1.00
Çok Zayıf	Zayıf	Orta	Güçlü	Çok Güçlü

## **H<sub>0</sub> – Sıfır Hipotezi (Null)**

“Fark yoktur” veya “Etki yoktur” iddiası.

Örnek: Yeni ilaç ile plasebo arasında fark yoktur.

Test sonucunda ya reddedilir ya da reddedilemez.

## **H<sub>1</sub> – Alternatif Hipotez**

“Fark vardır” veya “Etki vardır” iddiası.

Örnek: Yeni ilaç plaseboya göre daha etkilidir.

Kanıtlamak istediğimiz hipotezdir.

## **p-value (Anlamlılık Değeri)**

H<sub>0</sub> doğruyken gözlenen sonucu elde etme olasılığı.

- $p < 0.05$  → İstatistiksel olarak anlamlı
- $p < 0.01$  → Yüksek düzeyde anlamlı
- $p < 0.001$  → Çok yüksek düzeyde anlamlı

Düşük p-value = H<sub>0</sub>'a karşı güçlü kanıt

$$\alpha = 0.05$$

Anlamlılık Düzeyi  
(Eşik Değer)

- $\alpha = 0.05$  → %5 hata payı (en yaygın)
- $\alpha = 0.01$  → %1 hata payı (katı kriter)
- $p < \alpha$  ise H<sub>0</sub> reddedilir

## Hipotez Kur

## Veri Topla

## Test Seç

## Hesapla

## Karar Ver

1

$H_0$  ve  $H_1$ 'i net şekilde tanımla

2

Örneklem büyüklüğü ve yöntemi belirle

3

Veri türüne uygun istatistik testi seç

4

Test istatistiği ve p-value hesapla

5

$p < \alpha$  ise  $H_0$ 'ı reddet, sonucu yorumla

## Kritik Uyarılar

- İstatistiksel anlamlılık  $\neq$  pratik anlamlılık
- p-value etkinin büyüklüğünü göstermez
- Etki büyüklüğü (Cohen's d) mutlaka raporlayın
- Çoklu karşılaştırma? Bonferroni düzeltmesi
- Örneklem büyüklüğü yeterli mi? Güç analizi
- Varsayımları her zaman kontrol edin

## Bağımsız İki Örneklem T-Testi

İki grubun ortalamaları arasında istatistiksel olarak anlamlı bir fark olup olmadığını test eder.  
Önkoşul: Normal dağılım + varyans homojenliği

## T-Test Türleri

- 1 Bağımsız:** Farklı gruplar (kontrol vs deney)
- 2 Eşlenikli:** Aynı grup, önce-sonra ölçüm
- 3 Tek Örneklem:** Grup ortalaması vs bilinen değer

```
from scipy.stats import ttest_ind

# A grubu: Eski reklam    B grubu: Yeni reklam
grup_a = [23, 25, 28, 22, 30, 27, 24, 26, 29, 31]
grup_b = [30, 33, 35, 28, 38, 32, 36, 34, 31, 37]

t_stat, p_value = ttest_ind(grup_a, grup_b)

print(f"T istatistiği: {t_stat:.4f}")
print(f"p-value:      {p_value:.4f}")

if p_value < 0.05:
    print("✅ Gruplar arasında anlamlı fark var!")
```

## Tek Yönlü ANOVA (One-Way)

2 veya daha fazla grubun ortalamalarını aynı anda karşılaştırır.

- F istatistiği = Gruplar arası varyans / Grup içi varyans
- F büyük → Gruplar arası fark anlamlı
- Post-hoc test (Tukey) ile hangi gruplar farklı?

## T-Test vs ANOVA

- T-Test: 2 grup karşılaştırması
- ANOVA: 2+ grup karşılaştırması
- 3+ grup için çoklu t-test YAPMAYIN!  
(Tip I hata artar:  $\alpha' = 1-(1-\alpha)^k$ )
- ANOVA sonucu anlamlı → Post-hoc test
- Varsayım: Normal dağılım + eşit varyans

```
from scipy.stats import f_oneway

# Üç farklı öğretim yöntemiyle sınav puanları
yontem_a = [78, 82, 85, 79, 88, 90]
yontem_b = [65, 70, 68, 72, 66, 74]
yontem_c = [90, 92, 88, 95, 91, 87]

f_stat, p_value = f_oneway(yontem_a, yontem_b, yontem_c)
print(f"F istatistiği: {f_stat:.4f}, p-value: {p_value:.6f}")
```

## Ki-Kare Testi Nedir?

İki kategorik değişken arasında bağımsızlık olup olmadığını test eder.

Örn: Yolcu sınıfı ile hayatta kalma arasında ilişki var mı? (Titanic)

## Titanic Örneği: Çapraz Tablo

	Hayatta	Vefat
1. Sınıf	136	80
2. Sınıf	87	97
3. Sınıf	119	372

```
import pandas as pd
from scipy.stats import chi2_contingency

# Titanic veri seti
df = pd.read_csv('titanic.csv')
capraz = pd.crosstab(df['Pclass'], df['Survived'])

chi2, p_val, dof, expected = chi2_contingency(capraz)

print(f" $\chi^2 = {chi2:.2f}$ , p = {p_val:.6f}")
print(f"Serbestlik derecesi: {dof}")
# p < 0.05 → Sınıf ile hayatta kalma ilişkili!
```

## %95 Güven Aralığı Nedir?

Popülasyon parametresinin (gerçek ortalama) hangi aralıkta olduğuna dair tahmin.

- 100 kez örneklem alsan, 95'inde gerçek değer bu aralıkta olur
- Aralık dar → tahmin kesin
- Aralık geniş → belirsizlik yüksek

## Formül

$$\bar{X} \pm z \times (\sigma / \sqrt{n})$$

$\bar{X}$  = örneklem ortalaması

$z = 1.96$  (%95 için)

$\sigma$  = standart sapma,  $n$  = örneklem

## Örnek Hesaplama: Müşteri Memnuniyet Puanı

100

$n$  (veri sayısı)

7.2

$\bar{x}$  (ortalama)

1.5

$\sigma$  (std sapma)

1.96

$z$  (%95)

$$GA = 7.2 \pm 1.96 \times (1.5 / \sqrt{100}) = 7.2 \pm 0.294 = [6.906, 7.494]$$

## Kontrollü Deney Yöntemi

İki (veya daha fazla) versiyonun rastgele seçilmiş kullanıcı gruplarına sunulmasıyla performans farkının istatistiksel olarak ölçüldüğü deneysel yöntem. Veri odaklı karar vermenin temel taşıdır.



### Web Tasarımı

Buton rengi, sayfa düzeni,  
CTA metni, form uzunluğu



### E-Ticaret

Fiyatlandırma, ürün görseli,  
ödeme akışı, kampanya



### Mobil Uygulama

Onboarding akışı, bildirim  
zamanla, UI değişikliği



### Pazarlama

E-posta konu satırı, reklam  
metni, hedef kitle seçimi

## 1 Hedef Belirle

KPI seç: dönüşüm oranı,  
tıklama oranı, gelir...

## 2 Hipotez Kur

$H_0$ : Fark yok  
 $H_1$ : Yeni versiyon daha iyi

## 3 Örneklem Hesapla

Güç analizi ile gerekli  
kullanıcı sayısını bul

## 4 Uygula

Rastgele atama ile A/B  
gruplarına böl

## 5 Analiz Et

Yeterli veri sonrası  
istatistik testi uygula

## 6 Karar Ver

$p < \alpha$  ise kazananı  
kalıcı yap

## Senaryo

Mobil oyun şirketi yeni reklam formatını test ediyor. A: mevcut reklam (banner), B: yeni reklam (video). Metrik: günlük gelir (\$).

```
import numpy as np
from scipy.stats import ttest_ind

np.random.seed(42)
# A grubu: banner reklam (30 günlük gelir)
banner = np.random.normal(loc=45, scale=8, size=30)
# B grubu: video reklam (30 günlük gelir)
video = np.random.normal(loc=52, scale=10, size=30)

t_stat, p_value = ttest_ind(banner, video)

print(f"Banner ortalama: ${banner.mean():.2f}")
print(f"Video ortalama: ${video.mean():.2f}")
print(f"T-istatistik: {t_stat:.4f}")
print(f"p-value: {p_value:.4f}")

alpha = 0.05
if p_value < alpha:
    print("✅ Video reklam anlamlı fark yaratıyor!")
else:
    print("❌ Anlamlı fark bulunamadı")
```

Karar ↓ / Gerçek →	$H_0$ Doğru (Fark Yok)	$H_0$ Yanlış (Fark Var)
$H_0$ Reddedildi	<p><b>✗ Tip I Hata (<math>\alpha</math>)</b></p> <p>Fark yokken “fark var” demek. Örn: Etkisiz ilacı “etkili” ilan etmek. (Yanlış Pozitif)</p>	<p><b>✓ Doğru Karar</b></p> <p>Gerçekten fark varken “fark var” demek. (Doğru Pozitif — Güç = <math>1-\beta</math>)</p>
$H_0$ Reddedilmedi	<p><b>✓ Doğru Karar</b></p> <p>Gerçekten fark yokken “fark yok” demek. (Doğru Negatif)</p>	<p><b>⚠ Tip II Hata (<math>\beta</math>)</b></p> <p>Fark varken “fark yok” demek. Örn: Etkili ilacı “etkisiz” ilan etmek. (Yanlış Negatif)</p>

## istatistik\_temel.ipynb

Tanımlayıcı istatistik hesaplamaları, normal dağılım analizi, Shapiro-Wilk testi, korelasyon matrisi oluşturma

- 1 describe(), mean(), median()
- 2 Shapiro-Wilk normallik testi
- 3 Pearson & Spearman korelasyon
- 4 Histogram + Q-Q Plot

## ab\_testi.ipynb

A/B testi tasarımı, uygulaması ve sonuç analizi. T-test ile grup karşılaştırması, güven aralığı hesaplama

- 1 Rastgele grup ataması
- 2 ttest\_ind ile analiz
- 3 Güven aralığı hesaplama
- 4 Etki büyüklüğü (Cohen's d)

## chi\_square.ipynb

Titanic veri seti üzerinde Ki-Kare bağımsızlık testi, çapraz tablo analizi ve görselleştirme

- 1 pd.crosstab() kullanımı
- 2 chi2\_contingency testi
- 3 Heatmap görselleştirme
- 4 ANOVA ve post-hoc testler

## Haftalık Ödevler

1

### Ödev 1: Tanımlayıcı İstatistik

Bir veri seti için ortalama, medyan, mod, std hesaplayın ve normal dağılımı kontrol edin.

2

### Ödev 2: Korelasyon Analizi

En az 3 sayısal değişken arasında Pearson ve Spearman korelasyonu hesaplayın.

3

### Ödev 3: A/B Testi

Bir senaryo oluşturup t-test ile A/B testi gerçekleştirin ve güven aralığı hesaplayın.

4

### Ödev 4: Ki-Kare Testi

Titanic veri setinde farklı çapraz tablo kombinasyonları ile  $\chi^2$  testi uygulayın.

## Kaynaklar



### SciPy Stats Dokümantasyonu

[docs.scipy.org/doc/scipy/reference/stats.html](https://docs.scipy.org/doc/scipy/reference/stats.html)



### Think Stats (Allen Downey)

Ücretsiz e-kitap: [greenteapress.com/thinkstats](https://greenteapress.com/thinkstats)



### Khan Academy: İstatistik

Temel istatistik dersleri (Türkçe alt yazılı)



### Trustworthy Online Experiments

Kohavi et al. — A/B Testi referans kitabı



### Statistics for Data Science

Field, A. (2013) — *Discovering Statistics*

# Hafta 4 — Özet

1 Tanımlayıcı istatistik (ortalama, medyan, std) verinin hikayesini anlamamızı sağlar

2 Normallik testi, doğru istatistiksel yöntemi seçmek için önkoşuldur

3 Hipotez testleri (t-test, ANOVA,  $\chi^2$ ) farklı senaryolarda karar desteği sunar

4 A/B testi, veri odaklı karar vermenin en güçlü aracıdır

5 p-value tek başına yetmez: etki büyüklüğü ve güven aralığı da raporlayın

***“Veri olmadan fikir yürütmek, karşı takımı görmeden strateji kurmaya benzer.”***