

4. Hafta

İstatistiksel Deneyler ve Anlamlılık Testi

Kaynak: Bruce vd. (2026), Bölüm 3 — Statistical Experiments and Significance Testing

Doç. Dr. Onur Polat

Hacettepe Üniversitesi • Bilişim Enstitüsü

Tezsiz YL — Tüm Anabilim Dalları • Güz 2026–2027, Ders 4 / 14

GENEL BAKIŞ

Bu Haftanın Gündemi

Bölüm 3: İstatistiksel Deneyler

- A/B Testi ve kontrol grubu tasarımı
- Hipotez testleri: sıfır ve alternatif
- Yeniden örnekleme: permütasyon testi
- İstatistiksel anlamlılık ve p-değeri
- t-Testleri
- Çoklu test problemi
- ANOVA ve F-istatistiği
- Ki-kare testi

İleri Konular

- Serbestlik derecesi kavramı
- İki yönlü ANOVA
- Fisher'ın kesin testi
- Çok kollu haydut algoritması
- Güç ve örneklem büyüklüğü
- Veri bilimde p-değerinin rolü
- Python uygulamaları
- Epsilon-greedy ve Thompson örneklemesi

KLASİK İSTATİSTİKSEL ÇIKARIM

Deney Tasarımı İş Akışı



1. Hipotez Formüle Et:

"İlaç A mevcut standarttan daha iyi" veya "Fiyat B daha kârlı"

2. Deney Tasarla:

Hipotezi test edecek bir A/B testi veya deney düzenle

3. Veri Topla:

Deneyi uygula ve sonuçları kaydet

4. Çıkarım:

Sonuçları sınırlı örneklemden daha geniş popülasyona genelle

01

A/B Testi

İki tedaviyi karşılaştırarak hangisinin daha üstün olduğunu belirleme

A/B TESTİ: TEMEL TERİMLER

A/B Testi Temel Kavramları

Tedavi (Treatment)

Deneklerin maruz kaldığı şey (ilaç, fiyat, web başlığı)

Tedavi Grubu

Belirli bir tedaviye maruz kalan denekler grubu

Kontrol Grubu

Standart (veya hiç) tedavi uygulanan denekler grubu

Rastgeleleştirme

Deneklerin tedavilere rastgele atanması süreci

Denekler (Subjects)

Tedavilere maruz kalan birimler (web ziyaretçileri, hastalar vb.)

Test İstatistiği

Tedavinin etkisini ölçmek için kullanılan metrik

A/B TESTİ ÖRNEKLERİ

Web tasarımından tarıma kadar geniş kullanım alanı

- İki toprak işleminden hangisinin daha iyi tohum filizlenmesi sağladığını test etme
- İki tedaviden hangisinin kanseri daha etkili baskıladığını belirleme
- İki fiyattan hangisinin daha fazla net kâr getirdiğini test etme
- İki web başlığından hangisinin daha fazla tıklama aldığını ölçme
- İki web reklamından hangisinin daha fazla dönüşüm ürettiğini analiz etme

Tablo 3-1. E-ticaret deney sonuçları (2x2 tablo)

Sonuç	Fiyat A	Fiyat B
Dönüşüm	200	182
Dönüşüm Yok	23.539	22.406

NEDEN KONTROL GRUBU?

Kontrol grubu olmadan:

- "Diğer tüm koşullar eşit" güvencesi yoktur
- Fark gerçekten tedaviden mi yoksa şanstın mı kaynaklanıyor bilinmez

Kontrol grubu ile:

- Aynı koşullara tabi tutulur (tedavi hariç)
- Gözlenen her fark yalnızca iki kaynaktan gelebilir:
 1. Tedaviler arasındaki gerçek fark
 2. Rastgele atamadaki şans

Kör Çalışma (Blinding): Deneklerin hangi tedaviyi aldığını bilmediği çalışma. **Çift kör:** Araştırmacılar da bilmez. Bilişsel terapi gibi tedaviler için kör yapılamaz.

NEDEN SADECE A/B? NEDEN C, D DEĞİL?

A/B testleri popüler, ama tek istatistiksel deney türü değil!

Veri bilimcileri şu soruyla daha çok ilgilenir:

"Birden fazla olası fiyat arasından hangisi en iyi?"

yerine klasik soru:

"A ile B arasındaki fark istatistiksel olarak anlamlı mı?"

Çözüm: Çok Kollu Haydut (Multi-Arm Bandit) Algoritması

- Birden fazla tedaviyi eşzamanlı test eder
- Deney sırasında en iyi performansı gösterene yönelir
- *Detaylar bu bölümün ilerleyen kısımlarında*

02

Hipotez Testleri

Gözlenen etkinin şans eseri olup olmadığını sorgulama

HİPOTEZ TESTİ TERİMLERİ

Temel Kavramlar

Sıfır Hipotezi (H_0)

Şansın sorumlu olduğu hipotez — 'hiçbir fark yok'

Alternatif Hipotez (H_1)

Sıfır hipotezinin karşıtı — kanıtlamak istediğiniz şey

Tek Yönlü Test

Şans etkisini yalnızca bir yönde sayan hipotez testi

İki Yönlü Test

Şans etkisini her iki yönde de sayan hipotez testi

Neden hipotez testine ihtiyacımız var?

İnsanlar rastgele davranışlara karşı anlam atfetme eğilimindedir. İstatistiksel hipotez testleri, araştırmacıları rastgele şans tarafından kandırılmaktan korumak için geliştirilmiştir.

SIFIR VE ALTERNATİF HİPOTEZ

H_0 — Sıfır Hipotezi

"Tedaviler arasında fark yoktur; gözlenen her fark şansa bağlıdır."

Temel varsayım: Tedaviler eşdeğerdir ve aralarındaki her fark şanstıan kaynaklanır.

Amacımız bu hipotezi çürütmektir!

H_1 — Alternatif Hipotez

Örnekler:

- H_0 : "A ile B arasında fark yok"
 H_1 : "A, B'den farklı" (iki yönlü)
- H_0 : " $A \leq B$ "
 H_1 : " $A > B$ " (tek yönlü)
- H_0 : "B, A'dan X% daha iyi değil"
 H_1 : "B, A'dan X% daha iyi"

TEK YÖNLÜ vs İKİ YÖNLÜ TEST

Tek Yönlü (One-tail) Test

- Yeni seçenek (B), mevcut varsayılan (A) karşısında test edilir
- B kesin olarak daha iyi kanıtlanmadıkça A ile kalınır
- Yönlü (directional) alternatif hipotez: $H_1: B > A$
- Yalnızca bir yöndeki aşırı sonuçlar p-değerine sayılır

İki Yönlü (Two-tail) Test

- Her iki yönde de şanstıan kandırılma koruması istenir
- İki yönlü alternatif: $H_1: A \neq B$
- Her iki yöndeki aşırı sonuçlar p-değerine sayılır

Not: Python ve R varsayılan olarak iki yönlü test uygular. Veri bilminde tek vs iki yönlü ayrımı çok kritik değildir; p-değeri hassasiyeti bu düzeyde önemli olmayabilir.

03

Yeniden Örnekleme

Permütasyon testi ile rastgele varyasyonun rolünü keşfetme

YENİDEN ÖRNEKLEME TERİMLERİ

Temel Kavramlar

Permütasyon Testi

İki veya daha fazla örneklemini birleştirip rastgele yeniden örneklerle ayırma

Eşanlamlılar

Rastgeleleştirme testi, rastgele permütasyon testi, kesin test

Yeniden Örnekleme (Resampling)

Gözlenen veri setinden ek örnekler çekme

Yerine Koyarak / Koymadan

Bir öğenin çekildikten sonra havuza geri konulup konulmaması

Permütasyon Testi Prosedürü:

Grupları birleştir → Karıştır ve yeniden böl → İstatistiği hesapla → R kez tekrarla → Karşılaştır

PERMÜTASYON TESTİ ADIMLARI

- 1. Farklı grupların sonuçlarını tek bir veri setinde birleştir
- 2. Birleşik veriyi karıştır ve A grubu büyüklüğünde (yerine koymadan) rastgele çek
- 3. Kalan veriyi B grubu olarak ayır
- 4. C, D vb. gruplar varsa onlar için de aynısını yap
- 5. Orijinal örnekler için hesaplanan istatistiği, yeniden örnekler için de hesapla — bu bir permütasyon iterasyonudur
- 6. Adımları R kez tekrarlayarak test istatistiğinin permütasyon dağılımını elde et

Sonuç: Gözlenen fark permüte edilmiş farkların çoğundan büyükse → şans sorumlu değil → *istatistiksel olarak anlamlı*

ÖRNEK: WEB YAPIŞKANLIĞI

Sayfa A vs Sayfa B — Sayfada kalma süresi karşılaştırması

Senaryo: Yüksek değerli hizmet satan bir şirket, iki web sunumundan hangisinin daha iyi olduğunu test etmek istiyor. Satışlar nadir ve pahalı olduğu için proxy değişken olarak sayfada kalma süresi kullanılıyor.

$n_A = 21$

Sayfa A
görüntüleme sayısı

$n_B = 15$

Sayfa B
görüntüleme sayısı

35.67 sn

B – A ortalama fark
(B daha uzun)

%12.9

p-değeri
(anlamli DEĞİL)

Yorum: Permütasyon testi 1000 iterasyon sonucunda, rastgele farkların %12.9'u gözlenen farkı aştı. Bu fark şans varyasyonu aralığında → istatistiksel olarak anlamlı değil.

PYTHON: PERMÜTASYON TESTİ

Web yapışkanlığı örneği kodu

```
import numpy as np
import pandas as pd

session_times = pd.read_csv("web_page_data.csv")
session_times["Time"] *= 100

mean_a = session_times[session_times.Page == "Page A"].Time.mean()
mean_b = session_times[session_times.Page == "Page B"].Time.mean()

def perm_fun(x, n_a):
    x_resampled = np.random.permutation(x)
    x_a = x_resampled[:n_a]
    x_b = x_resampled[n_a:]
    return x_b.mean() - x_a.mean()

n_a = session_times[session_times.Page == "Page A"].shape[0]
perm_diffs = [perm_fun(session_times.Time, n_a)
               for _ in range(1000)]

p_value = np.mean(perm_diffs > mean_b - mean_a)
print(f"p-değeri: {p_value:.4f}")
```

PERMÜTASYON TÜRLERİ

Rastgele Permütasyon

Verileri rastgele karıştırarak R kez tekrarlama. En yaygın yaklaşım.

Tüketici (Exhaustive)

Tüm olası bölünmeleri hesaplama. Küçük örneklem için pratik. Kesin test olarak da bilinir.

Bootstrap Permütasyon

Yerine koyarak örnekleme. Hem tedavi ataması hem de popülasyon seçimi rastgeleliğini modeller.

Veri Bilimi İçin Sonuç: Permütasyon testleri kodlaması kolay, yorumlaması kolay ve formül tabanlı yöntemlere pratik bir alternatiftir. Veri sayısal veya ikili olabilir; normal dağılım varsayımı gerekmez.

04

İstatistiksel Anlamlılık ve p-Değeri

Sonucun şans eseri olup olmadığını ölçme

p-DEĞERİ VE ALPHA

Temel Kavramlar

p-değeri

Sıfır hipotezi doğruyken, gözlenen sonuç kadar aşırı veya daha aşırı bir sonuç elde etme olasılığı

Alpha (α)

İstatistiksel anlamlılık eşiği. Tipik değerler: %5 ve %1

Tip 1 Hata

Gerçek olmayan bir etkiyi gerçek sanmak (yanlış pozitif)

Tip 2 Hata

Gerçek bir etkiyi tespit edememek (yanlış negatif)

Dikkat: p-değeri "sonucun şansa bağlı olma olasılığı" DEĞİLDİR. Doğrusu: "Bir şans modeli verildiğinde, gözlenen kadar aşırı sonuçların olma olasılığı" — ince ama önemli bir fark.

ÖRNEK: FİYAT TESTİ

E-ticaret dönüşüm oranları karşılaştırması

Sonuç	Fiyat A	Fiyat B
Dönüşüm	200	182
Dönüşüm Yok	23.539	22.406

%0.842

5

Fiyat A
dönüşüm oranı

%0.805

7

Fiyat B
dönüşüm oranı

0.0368

Yüzde puan farkı
(A lehine)

p ≈ 0.31

p-değeri
(anlamli DEĞİL)

Yorum: 45.000'den fazla veri noktası olmasına rağmen, dönüşüm sayıları yalnızca yüzlerce. $p \approx 0.31$ ile fark şans varyasyonu dahilinde → istatistiksel olarak anlamlı değil.

TİP 1 VE TİP 2 HATALAR

	H_0 Gerçekten Doğru	H_0 Gerçekten Yanlış
H_0 Reddedilmedi	Doğru karar (Gerçek negatif)	TİP 1 HATA (Yanlış pozitif)
H_0 Reddedildi	TİP 2 HATA (Yanlış negatif)	Doğru karar (Gerçek pozitif)

Hipotez testlerinin temel amacı Tip 1 hatalardan korumaktır. Tip 2 hata aslında hata değil, örneklem büyüklüğünün etkiyi tespit etmeye yeterli olmadığını gösterir.

VERİ BİLİMİNDE p-DEĞERİ

p-Değeri Tartışması

- Bazı psikoloji dergileri p-değerine dayalı yayın kararlarını yasakladı
- 2016'da ASA 6 ilke yayınladı: p-değeri tek başına yeterli kanıt değil
- İstatistiksel anlamlılık $\neq$ pratik anlamlılık

Veri Bilimciler İçin Pragmatik Bakış

- p-değeri bir model sonucunun normal şans varyasyonu içinde olup olmadığını belirlemede faydalı bir metriktir
- Kontrol eden değil, kararı bilgilendiren bir araç olarak kullanılmalı
- Makine öğrenmesinde: bir değişkenin modele dahil edilip edilmeyeceğine karar vermede ara girdi olarak kullanılabilir

t-TESTLERİ

Temel Kavramlar

Test İstatistiği

İlgilenilen fark veya etki için bir metrik

t-istatistiği

Ortalamalar gibi yaygın test istatistiklerinin standartlaştırılmış biçimi

t-dağılımı

Sıfır hipotezinden türetilen referans dağılım; gözlenen t-istatistiği bununla karşılaştırılır

```
from scipy import stats

res = stats.ttest_ind(
    session_times[session_times.Page == "Page A"].Time,
    session_times[session_times.Page == "Page B"].Time,
    equal_var=False)
print(f"p-değeri (tek yönlü): {res.pvalue / 2:.4f}")
# Çıktı: p-değeri (tek yönlü): 0.1408
```

ÇOKLU TEST PROBLEMİ

Alpha enflasyonu ve yanlış keşif oranı

Temel Kavramlar

Tip 1 Hata

Bir etkinin istatistiksel olarak anlamlı olduğu sonucuna yanlışlıkla varmak

Yanlış Keşif Oranı (FDR)

Birden fazla testte Tip 1 hata yapma oranı

Alpha Enflasyonu

Test sayısı arttıkça Tip 1 hata olasılığının artması

p-değeri Düzeltmesi

Aynı veri üzerinde çoklu test yapılmasını hesaba katma

Aşırı Uyum (Overfitting)

Gürültüye uyum sağlama

Örnek: 20 rastgele değişken + 1 sonuç $\rightarrow \alpha=0.05$ 'te en az birinin yanlışlıkla anlamlı çıkma olasılığı: $1 - 0.95^{20} = \%64$

05

ANOVA

Birden fazla grup arasındaki farkları analiz etme

ANOVA: TEMEL KAVRAMLAR

Varyans Analizi Terimleri

Eşleşik Karşılaştırma (Pairwise)

Birden fazla grup arasında ikişerli hipotez testi

Omnibus Test

Birden fazla grup ortalaması arasında genel varyans testi

Varyans Ayırıştırma

Bireysel değere katkıda bulunan bileşenleri ayırma

F-istatistiği

Grup ortalamaları arasındaki farkın bir şans modelinde beklenenin üzerinde olup olmadığını ölçen standart istatistik

SS (Kareler Toplamı)

Bir ortalama değerden sapmaların kareleri toplamı

ANOVA = A/B testinin çoklu gruplara genelleştirilmesidir. İkili karşılaştırmalar yerine tek bir genel test yapılır.

ÖRNEK: DÖRT WEB SAYFASI

A/B/C/D testi — yapışkanlık (saniye)

	Sayfa 1	Sayfa 2	Sayfa 3	Sayfa 4
	164	178	175	155
	172	191	193	166
	177	182	171	164
	156	185	163	170
	195	177	176	168
Ortalama	172	185	176	162

Genel Ortalama: 173.75

4 grup → 6 eşleşik karşılaştırma. **Çoklu karşılaştırma sorunu!** ANOVA tek bir omnibus test ile tüm grupları birden test eder.

```
import statsmodels.api as sm
import statsmodels.formula.api as smf

model = smf.ols("Time ~ Page", data=four_sessions).fit()
```

```
sm.stats.anova_lm(model)
# F=2.74, p=0.0776 (anlamlılık eşiğinin altında)
```

F-İSTATİSTİĞİ VE VARYANS AYRIŞTIRMA

Varyans Ayrıştırma (Decomposition of Variance)

Her gözlem = Genel Ortalama + Tedavi Etkisi + Artık Hata

Örnek (sol üst hücre = 164):

1. Genel ortalama: 173.75
2. Tedavi etkisi: -1.75 (172 - 173.75)
3. Artık: -8 (164 - 172)
4. **Toplam: 164 ✓**

	df	SS	MS	F	p
Sayfa	3	831.4	277.1	2.74	0.0776
Artık	16	1618.4	101.2		

$F = MS(\text{Tedavi}) / MS(\text{Hata}) = 277.1 / 101.2 = 2.74$

İKİ YÖNLÜ ANOVA

Tek Yönlü ANOVA: Yalnızca bir faktör değişiyor (örn: sayfa tasarımı)

İki Yönlü ANOVA: İki faktör birden değişiyor (örn: sayfa tasarımı + hafta sonu/hafta içi)

İki yönlü ANOVA'da 3 etki test edilir:

1. Ana etki — Faktör A (sayfa tasarımı)
2. Ana etki — Faktör B (hafta sonu/hafta içi)
3. Etkileşim etkisi ($A \times B$)

ANOVA, regresyon ve lojistik regresyona giden yolda ilk adımdır — birden fazla faktör ve etkileri modellenenebilir.

06

Ki-Kare Testi

Sayım verisiyle beklentiden sapmanın anlamlılığını ölçme

Kİ-KARE: BAŞLIK TESTİ ÖRNEĞİ

Üç farklı başlığın tıklama oranı karşılaştırması

	Başlık A	Başlık B	Başlık C
Tıklama	14	8	12
Tıklama Yok	986	992	988

Beklenen (H_0 : eşit tıklama oranı): $34/3 = 11.33$ tıklama / başlık

Pearson Artığı (R) = $(\text{Gözlenen} - \text{Beklenen}) / \sqrt{\text{Beklenen}}$

Ki-kare istatistiği = $\sum R^2 = 1.666$

Serbestlik derecesi = $(r-1) \times (c-1) = 1 \times 2 = 2$

p-değeri $\approx 0.4348 \rightarrow$ **Anlamlı DEĞİL — fark şans eseri olabilir**

```
from scipy import stats
survivors = np.array([[14, 8, 12], [986, 992, 988]])
res = stats.chi2_contingency(survivors)
print(f"Ki-kare: {res.statistic:.4f}, p: {res.pvalue:.4f}")
```

FISHER'IN KESİN TESTİ

Ne zaman kullanılır?

- Sayılar çok küçük olduğunda (özellikle tek haneli)
- Ki-kare yaklaşımı yetersiz kaldığında
- Tüm olası permütasyonları numaralayıp kesin p-değeri hesaplar

Nasıl çalışır?

- Verinin oluşabilecek tüm olası düzenlemelerini hesaplar
- Her düzenlemenin olasılığını tablolayarak kesin p-değerini verir
- Başlık örneği: Fisher $p \approx 0.4824$ (yeniden örnekleme $p \approx 0.4853$)

Veri Bilimi İçin: Ki-kare testi, özellikle örneklem büyüklüğü hesaplamalarında ve web deneylerinde kullanışlıdır. Otomatik özellik seçiminde de sınıf prevalansının rastgele varyasyon dışında olup olmadığını test edebilir.

07

Çok Kollu Haydut Algoritması

Keşif ve sömürü arasında optimum denge

ÇOK KOLLU HAYDUT ALGORİTMASI

Geleneksel A/B testine modern alternatif

Geleneksel A/B Testi

- Sabit örneklem dağılımı
- Sonuç: "etki kanıtlanmadı"
- Düşük performanslı tedaviye fazla maruz kalma
- Katı ve esnek olmayan süreç
- 2+ tedavi için karmaşık

Çok Kollu Haydut

- Dinamik örneklem dağılımı
- Deney sırasında en iyiye yönelir
- Düşük performanslıya daha az maruz kalma
- Esnek ve uyarlanabilir
- 3+ tedaviyi doğal olarak destekler

Slot makinesi analojisi: Her kol farklı oranda ödeme yapar. Amaç: toplam kazancı maksimize etmek — en kazançlı kolu mümkün olduğunca erken bulmak.

EPSİLON-GREEDY ALGORİTMASI

A/B testi için basit bir epsilon-greedy algoritması:

1. $[0, 1]$ arasında düzgün dağılımlı rastgele sayı üret
2. Sayı $[0, \epsilon]$ arasındaysa $\rightarrow$ keşif modunda yazı-tura at:
 - a. Yazı $\rightarrow$ A teklifini göster, b. Tura $\rightarrow$ B teklifini göster
3. Sayı $\geq \epsilon \rightarrow$ sömürü modunda, o ana kadar en yüksek yanıt oranına sahip teklifi göster

$$\epsilon = 1$$

Tam rastgele
(standart A/B)

$$\epsilon = 0$$

Tam açgözlü
(keşif yok)

$$0 < \epsilon < 1$$

Hibrit yaklaşım
(optimum)

İleri: Thompson örneklemesi (Bayesci yaklaşım) — beta dağılımı ile prior güncelleme, her çekimde en iyi kolu seçme olasılığını maksimize eder.

GÜÇ VE ÖRNEKLEM BÜYÜKLÜĞÜ

Temel Kavramlar

Etki Büyüklüğü (Effect Size)

Tespit etmeyi umduğunuz minimum etki (örn: tıklama oranında %20 artış)

Güç (Power)

Belirli bir örneklem büyüklüğü ile verilen etki büyüklüğünü tespit etme olasılığı

Anlamlılık Düzeyi (α)

Testin uygulanacağı istatistiksel anlamlılık seviyesi

Güç hesaplamasında 4 birbirine bağlı bileşen:

n

Örneklem
Büyüklüğü

Örnek

anlamlılık

d

Etki
Büyüklüğü

α

Anlamlılık
Düzeyi

$1-\beta$

Güç
(Power)

ÖRNEKLEM BÜYÜKLÜĞÜ HESABI

Reklam tıklama oranı örneği

Senaryo: Mevcut tıklama oranı %1.1, yeni reklamın %10 daha iyi yapmasını (%1.21) istiyoruz.

- $n=300$, %10 fark → **tespit edemez**
- $n=2000$, %50 fark → **yine tespit edemez (hâlâ binlerce gösterim gerekli)**

```
import statsmodels.api as sm

# Etki büyüklüğü hesaplama
effect_size = sm.stats.proportion_effectsize(0.0121, 0.011)

# Örneklem büyüklüğü hesaplama
analysis = sm.stats.TTestIndPower()
result = analysis.solve_power(
    effect_size=effect_size,
    alpha=0.05, power=0.8, alternative="larger")
print(f"Gereken örneklem: {result:.0f}")
# Çıktı: 116.602 (her gruptan 120K gösterim)
```

SERBESTLİK DERECEŚİ

İstatistiksel testlerin ve dağılımların anahtar parametresi

Serbestlik derecesi (d.f.) nedir?

Örneklem verisinden hesaplanan istatistiklere uygulanan bir kavramdır ve serbestçe değişebilen değer sayısını ifade eder.

Örnek: 10 değer in ortalamasını biliyorsanız, 9 değeri serbestçe seçebilirsiniz — 10. değeri otomatik belirlenir → d.f. = 9

Neden önemli?

- Varyans ve std hesabında n yerine $n-1$ kullanılmasının nedeni (yanlılık düzeltmesi)
- t-dağılımı, F-dağılımı, ki-kare dağılımı şekillerini belirler
- Regresyonda: kategorik değişkenleri $n-1$ gösterge (dummy) ile kodlama (çoklu doğrusal bağlantıdan kaçınma)

Veri bilimi için: Büyük veri setlerinde n vs $n-1$ farkı ihmal edilebilir. Daha çok regresyon ve gösterge kodlamada önemlidir.

BÖLÜM ÖZETİ

Deney Tasarımı

Rastgeleleştirme ile kontrol ve tedavi gruplarına ayırma — geçerli çıkarımın temeli

Hipotez Testleri

H_0 (şans) vs H_1 (gerçek etki) — istatistiksel çıkarım çerçevesi

Yeniden Örneklem

Permütasyon testleri: kodlaması ve yorumlaması kolay, formül gerektirmez

p-Değeri

Pragmatik bir araç — kararı bilgilendiren, kontrol etmeyen bir metrik

ANOVA & Ki-Kare

Çoklu grup ve sayım verisi karşılaştırmaları

Çok Kollu Haydut

Keşif/sömürü dengesi — geleneksel A/B testine esnek alternatif