Hipotez Testlerinin Türleri

Artık hipotez testine NEDEN ihtiyacımız olduğunu ve arkasındaki temel mantığı anladığınıza göre, istatistiksel alet çantamızdaki farklı araçlardan bahsedelim. Bunu farklı çekiç türlerini öğrenmek gibi düşünün - bir resim çerçevesini asmak için balyoz kullanmazsınız, değil mi? Benzer şekilde, farklı istatistiksel durumlar farklı testler gerektirir.

Ancak yeni başlayanların en çok kafasını karıştıran şey şu: O kadar çok farklı test var ki! Hangisini kullanacağımı nasıl bileceğim?

Endişelenmeyin. Bu bölümün sonunda, her testin ne işe yaradığını değil, daha da önemlisi, NE ZAMAN ve NEDEN kullanılması gerektiğini anlayacaksınız. Ayrıca, insanların aklını karıştıran en yaygın kafa karışıklıklarını da açıklığa kavuşturacaksınız.

Parametric vs Non-Parametric Testler

Spesifik testlere dalmadan önce, hipotez test dünyasındaki temel bir ayrımı anlayalım.

Fark Nedir?

Diyelim ki öğrencilerin performansını değerlendiren bir öğretmensiniz.

Senaryo A - Parametrik Yaklaşım: Öğrencilere 0-100 puanlı bir matematik testi veriyorsunuz. Ortalama puanı (75), standart sapmayı (12) hesaplıyor ve puanların çan eğrisi düzenini (normal dağılım) izlediğini varsayıyorsunuz. "Öğrencilerin çoğu 63 ile 87 arasında puan aldı" gibi ifadeler kullanabilirsiniz.

Senaryo B - Parametrik Olmayan Yaklaşım: Öğrencilerden en sevdikleri dersleri 1'den 5'e kadar sıralamalarını istiyorsunuz. Öğrenci A, Matematik'in 1 numara olduğunu, Öğrenci B ise Matematik'in 4 numara olduğunu söylüyor. "Ortalama sıralamayı" anlamlı bir şekilde hesaplayamazsınız ("ortalama 2.7" ne anlama geliyor ki?). Burada kesinlikle bir çan eğrisi yok. Bu yüzden farklı bir yaklaşım kullanıyorsunuz - belki de kaç öğrencinin Matematik'i ilk 3'üne koyduğunu sayıyorsunuz.

Temel Fark:

Parametreli testler verileriniz hakkında varsayımlarda bulunur:

- Verileriniz belirli bir deseni (genellikle normal bir dağılım - çan eğrisi şekli) takip eder
- Ortalaması çıkarılabilecek ölçümlerle çalışıyorsunuz (boy, kilo, test puanları gibi)
- Anlamlı bir şekilde ortalamaları ve standart sapmaları hesaplayabilirsiniz

Parametrik olmayan testler şu varsayımları yapmaz:

- Verileriniz belirli bir modeli takip etmeyebilir
- Sıralamalar, kategoriler veya sıralı verilerle çalışıyor olabilirsiniz
- Ortalamayı hesaplayamazsınız veya hesaplamamalısınız

Gerçek Dünyadan Örnekler

Parametrik Durum

- Öğrencilerin boyları (boyların ortalamasını alabilirsiniz, kabaca bir çan eğrisi izlerler)
- Bir görevi tamamlamak için gereken süre (anlamlıdan ortalamaya, genellikle normale yakın)
- Kan basıncı ölçümleri (sayısal, sürekli, genellikle normal dağılımlı)

Parametrik Olmayan Durum

- Müşteri memnuniyeti derecelendirmeleri (1-5 yıldız) - 5 gerçekten 1'den "5 kat daha mı iyi"? Pek sayılmaz.
- Tıbbi durumun şiddeti (hafif, orta, şiddetli) - "orta + şiddetli = ??" ortalamasını alamazsınız.
- Favori renk seçimleri (kırmızı, mavi, yeşil) - tamamen kategorik

Yaygın Bir Karışıklık: "Ama benim verilerim sayılar içeriyor, dolayısıyla parametrik, değil mi?"

YANLIŞ! Sayılara sahip olmanız, parametrik testlerin uygun olduğu anlamına gelmez.

Örnek: Müşterilerinize "Ne kadar memnunsunuz?" sorusunu şu seçeneklerle soruyorsunuz:

- 1 = Hiç Memnun Değilim
- 2 = Memnun Değilim
- 3 = Nötrüm
- 4 = Memnunum
- 5 = Çok Memnunum

Evet, sayılarınız var (1-5). Ama kendinize şunu sorun:

- "Hiç Memnun Değilim" ile "Memnun Değilim" arasındaki fark, "Memnunum" ile "Çok Memnunum" arasındaki farkla aynı mıdır? (Muhtemelen hayır!)
- Ortalama memnuniyetin 3.7 olduğunu söylemek mantıklı mı? (Bir bakıma, ama "ortalama boy 170 cm'dir" demek kadar anlamlı değil)
- Memnuniyet derecelendirmeleri çan eğrisini mi takip ediyor? (Genellikle hayır - çarpık olabilirler)

Bunlar sıralı(ordinal) verilerdir; sayılar sırayı temsil eder, ancak aralarındaki boşluklar eşit olmayabilir. Bu durumda parametrik olmayan testler genellikle daha iyidir.

Pratik Kural:

- Parametrik kullanımı şu durumlarda gereklidir: Sürekli ölçümler, veriler yaklaşık olarak çan şeklinde görünüyorsa, iyi örneklem büyüklüğü (genellikle $n \geq 30$)
- Şu durumlarda parametrik olmayanı kullanın: Kategorik/sıralı veriler, çok çarpık veriler, küçük örnekler veya parametrik varsayımların ciddi şekilde ihlal edildiği durumlar

Burada, daha yaygın olarak öğretildiği ve kullanıldığı için esas olarak parametrik testlere odaklanılacaktır, ancak parametrik olmayan alternatiflerin mevcut olduğu durumlarda da belirtilecektir.

Test Seçim Çerçevesi: Neleri Karşılaştırıyorsunuz? Doğru testi seçmenin sırrı şudur: Neyi karşılaştırdığınızı kendinize sorun. Şunları mı karşılaştırıyorsunuz:

1. Bir grubu bilinen bir değere mi dönüştürüyorum? (Örneklemim beklentilerimden farklı mı?)
2. Birbirine göre iki grup mu? (Bu iki grup farklı mı?)
3. İki gruptan fazla mı? (Bu gruplardan herhangi biri birbirinden farklı mı?)
4. Kategoriler arasındaki ilişkiler? (Bu iki şey birbiriyle ilişkili mi?)

Şimdi size mantıklı gelecek bir karar ağacı çizelim:

Başlangıç: Neyi karşılaştırıyorum?

├─ Bilinen bir değere tek bir örnek │ ├─ Ortalamayı mı test ediyorsunuz? → One-sample t-test or z-test │ └─ Bir oranı mı test ediyorsunuz? → One-sample proportion test │ ├─ İki Grup │ ├─ Aynı kişiler iki kez ölçüldü mü? → Paired t-test │ ├─ Her grupta farklı insanlar mı var? → Two-sample t-test │ └─ Oranları mı karşılaştırıyorsunuz? → Two-proportion z-test │ ├─ Üç veya daha fazla grup (ortalamaları karşılaştırarak) │ └─ → ANOVA (Analysis of Variance) │ └─ Kategoriler/frekans sayıları └─ → Chi-square test

Z Testi: Popülasyon Standart Sapmasını Bildiğinizde

Z-Testi Nedir?

Z-testi şu durumlarda kullanılır:

1. Bir örnek ortalamasını bir popülasyon ortalamasıyla karşılaştırıyorsunuz
2. Popülasyon standart sapmasını (σ) biliyorsunuz
3. VEYA örneğiniz büyükse ($n \geq 30$)

Mesele şu ki: Gerçek hayatta σ'yi (popülasyon standart sapmasını) neredeyse hiç bilemezsiniz.

Bir düşünün: Popülasyon standart sapmasını biliyorsanız, bu zaten tüm nüfusu ölçtüğünüz anlamına gelir! Peki bunu yaptıysanız, neden bir örneklem üzerinde hipotez testi yapıyorsunuz?

Peki z-testlerini ne zaman kullanırız?

Gerçekçi Senaryo 1 - Geçmiş Veriler: 10 yıldır civata üreten bir fabrika işletiyorsunuz. Milyonlarca cıvatadan, popülasyon standart sapmasının $\sigma = 0{,}05$ mm olduğunu biliyorsunuz. Şimdi yeni bir makine kuruyorsunuz. Yeni makineden 50 civatalık bir örneklem alarak ortalama çapın hala 5 mm olup olmadığını kontrol ediyorsunuz. Burada, geçmiş verilerden σ değerini bildiğiniz için bir z testi kullanabilirsiniz.

Gerçekçi Senaryo 2 - Büyük Örneklem: σ'yi bilmeseniz bile, örneğiniz yeterince büyükse ($n \geq 30$), örneklem standart sapması (s), σ'nin iyi bir tahminidir ve z-testini kullanabilirsiniz. Bu, Merkezi Limit Teoremi (temel olarak şunu söyler: herhangi bir dağılımdan büyük örnekler normal görünmeye başlar) sayesinde işe yarar.

Z-Test Türleri

1. One-Sample Z-Test

Cevapladığı soru: "Örneklem ortalaması, bilinen bir popülasyon ortalamasından önemli ölçüde farklı mı?"

Formül

z = $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$\bar{x}$ = örneklem ortalaması μ_0 = varsayılan popülasyon ortalaması σ = popülasyon standart sapması n = örneklem boyutu

Bu formülün gerçek anlamı şudur:

- $(\bar{x} - \mu_0)$: Bu, "Örneklem ortalamasının beklentilerimden ne kadar uzakta olduğu?" sorusudur.
- $(\sigma/\sqrt{n})$: Bu, "Rastgele şansa bağlı olarak ne kadar varyasyon beklemeliyim?" sorusudur (standart hata olarak adlandırılır)
- z :"Sonucum beklenenden kaç 'standart hata' uzakta?" sorusudur.

Bunu şöyle düşünün: Bir şeyi cetvelle ölçüyorsanız, z-puanı size sonucun beklediğinizden "kaç cetvel uzunluğu uzakta" olduğunu söyler.

Tam Çözümlü Örnek:

Senaryo: Ulusal standart bir sınavın bilinen ortalama puanı μ = 500 ve standart sapması σ = 100'dür (yıllar içinde milyonlarca sınav katılımcısına göre). Yeni bir hazırlık kursu, puanları iyileştirdiğini iddia ediyor. Hazırlık kursunu alan 40 öğrenciyi test ediyorsunuz.

Adım 1: Hipotezleri kurun

- H_0: $\mu \leq 500$ (Hazırlık kursu yardımcı olmuyor)
- H_a: $\mu > 500$ (The prep course helps)

Adım 2: Verileri toplayın

- Örneklem büyüklüğü: n = 40 öğrenci
- Örneklem ortalaması: $\bar{x}$ = 525
- Nüfus standart sapması: σ = 100 (tarihsel verilerden bilinmektedir)

Adım 3: Test istatistiğini hesaplayın z = $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ z = (525 - 500) / (100 / √40) z = 25 / (100 / 6.32) z = 25 / 15.81 z = 1.58

Adım 4: Z puanını yorumlayın 1,58'lik bir z-puanı şu anlama gelir: "Örneklem ortalaması, H_0 doğru olsaydı beklediğimizden 1,58 standart hata daha yüksektir."

z = 1.58 bize ne anlatıyor?

H_0 doğru olsaydı (hazırlık kursu işe yaramasaydı), tüm olası örneklem ortalamalarının dağılımının bu olduğunu varsayalım. Örneklem ortalamalarının çoğu 500 civarında kümelenirdi, ancak bazıları şans eseri daha yüksek veya daha düşük olurdu. 1,58'lik bir z puanı, örneklem ortalamamızın (525) bu dağılımın üst kısmında olduğunu, ancak çok da uzakta olmadığını gösterir.

Adım 5: Karar verin

z = 1,58'in H_0'yi reddetmek için "yeterince aşırı" olup olmadığına karar vermek için kritik değere veya p-değerine bakıyoruz (bunu bir sonraki derste ayrıntılı olarak ele alacağız).

α = 0,05'te (yüzde 95 güven) sağ kuyruklu bir test için kritik z değeri 1,645'tir.

z = 1,58 < 1,645 olduğundan H_0'yi reddedemeyiz.

Sonuç: "Bu örneğe dayanarak, hazırlık kursunun sınav puanlarını iyileştirdiği sonucuna varmak için yeterli güçlü kanıt yok. Gözlemlediğimiz 25 puanlık artışın tesadüfen gerçekleşmesi makul olabilir."

Önemli not: Bu, kursun işe yaramadığını kanıtlamıyor! Sadece elimizdeki kanıtların işe yaradığını iddia edecek kadar güçlü olmadığı anlamına geliyor.

In [8]:
```
#from statsmodels.stats.weightstats import ztest
# veriler varsa
from math import sqrt
from scipy.stats import norm

n = 40
x_bar = 525
sigma = 100
mu0 = 500

# Z istatistiği
z = (x_bar - mu0) / (sigma / sqrt(n))

# Tek kuyruklu p-değerleri
# H1: ortalama > mu0 ise (sağ kuyruk)
p_right = 1 - norm.cdf(z)

# H1: ortalama < mu0 ise (sol kuyruk)
p_left = norm.cdf(z)

# p-değeri (2 kuyruklu test)
p_twotail = 2 * (1 - norm.cdf(abs(z)))

print(f"Z istatistiği: {z:.3f}")
print(f"Sağ kuyruk p-değeri (H1: >): {p_right:.5f}")
print(f"Sol kuyruk p-değeri (H1: <): {p_left:.5f}")
print(f"İki kuyruklu p-değeri: {p_twotail:.5f}")
```

```
Z istatistiği: 1.581
Sağ kuyruk p-değeri (H1: >): 0.05692
Sol kuyruk p-değeri (H1: <): 0.94308
İki kuyruklu p-değeri: 0.11385
```

2. Two-Sample Z-Test

Cevapladığı soru: "İki farklı popülasyonun ortalamaları anlamlı derecede farklı mıdır?"

Formül

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\bar{x}_1, \bar{x}_2$: grup 1 ve grup 2'nin örnek ortalamaları σ_1, σ_2 = popülasyon standart sapmaları (bilinen) n_1, n_2 = örneklem büyüklükleri

Örnek

Senaryo: Erkek ve kadın yazılım mühendislerinin ortalama maaşlarını karşılaştırmak istiyorsunuz. Sektör genelindeki verilerden şunları biliyorsunuz:

Erkek mühendisler: σ_1 = 50.000 Kadın mühendisler: σ_2 = 45.000

Anketiniz 60 erkek mühendis: $\bar{x}_1$ = 850000, n_1 = 60 55 kadın mühendis: $\bar{x}_2$ = 820000, n_2 = 55

Hipotezler:

H_0: $\mu_1 = \mu_2$ (Maaş farkı yok) H_a: $\mu_1 \neq \mu_2$ (Maaş farkı var)

Bu iki yönlü bir testtir çünkü herhangi bir fark olup olmadığını (her iki yönde de) kontrol ediyoruz.

z'yi hesaplayalım:

z = (850000 - 820000) / √(50000^2/60 + 45000^2/55) z = 3000 / √(41666666.66 + 36818181.81) z = 30000 / 8859.16 z = 3.38

3,39'luk bir z puanı oldukça uç bir değerdir! α = 0,05'te iki taraflı bir test için kritik değerler ±1,96'dır.

3.39 > 1.96 olduğundan H_0'yi reddediyoruz.

In [17]:

```python
from math import sqrt
from scipy.stats import norm

# Veriler
x1_bar = 850000     # 1. örneklem ortalaması
x2_bar = 820000     # 2. örneklem ortalaması
sigma1 = 50000      # 1. grubun popülasyon std sapması
sigma2 = 45000      # 2. grubun popülasyon std sapması
n1 = 60
n2 = 55
```

```
# Z istatistiği
z = (x1_bar - x2_bar) / sqrt((sigma1**2 / n1) + (sigma2**2 / n2))

# p-değeri (çift kuyruklu örnek)
p = 2 * (1 - norm.cdf(abs(z)))

print(f"Z istatistiği: {z:.3f}")
print(f"p-değeri: {p:.5f}")
```

```
Z istatistiği: 3.386
p-değeri: 0.00071
```

3. Proportion Z-Test

Cevapladığı soru: "Örneklemdeki oran, bilinen bir orandan önemli ölçüde farklı mı?"

Formül

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$\hat{p}$: Örneklem oranı p_0 : varsayılan(hipotez edilen) popülasyon oranı n : örneklem büyüklüğü

Örnek

Senaryo: Bir şirket, müşterilerinin %90'ının memnun olduğunu iddia ediyor. Siz ise bunun daha düşük olduğunu düşünüyorsunuz. 200 rastgele müşteriyle anket yapıyorsunuz ve 170'inin memnun olduğunu görüyorsunuz.

- Adım 1 : Örnek oranını hesapla $\hat{p}$ = 170 / 200 = 0.85 (%85)
- Adım 2 : Hipotezleri Kur
  - H_0: p ≥ 0.90
  - H_a: p < 0.90

  Bu sol kuyruklu bir testtir.
- z hesaplanır. z = (0.85 - 0.90) / √(0.90 × 0.10 / 200) z = -0.05 / √(0.09 / 200) z = -0.05 / √0.00045 z = -0.05 / 0.0212 z = -2.36
- Adım 4: Karar α = 0,05'te sol kuyruklu bir test için kritik değer -1,645'tir.

-2.36 < -1.645 (daha uç) olduğundan H_0'yi reddediyoruz.
Sonuç: "Gerçek memnuniyet oranı muhtemelen %90'ın altındadır. Şirketin iddiası abartılı görünüyor."

In [1]:
```
from math import sqrt
from scipy.stats import norm

# Verilen değerler
n = 200                    # örneklem büyüklüğü
x = 170                    # memnun müşteri sayısı
p_hat = x / n              # örneklem oranı
p0 = 0.90                  # varsayılan (hipotez edilen) oran

# Z istatistiği hesaplama
```

```python
z = (p_hat - p0) / sqrt((p0 * (1 - p0)) / n)

# p-değeri (tek kuyruklu, H1: p < p0)
p_value = norm.cdf(z)

# Sonuçları yazdır
print(f"Z istatistiği: {z:.3f}")
print(f"p-değeri: {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: Müşteri memnuniyet oranı %90'dan düşüktür.")
else:
    print("H0 reddedilemez: Veriler %90 memnuniyet oranına aykırı değildir.")
```

```
Z istatistiği: -2.357
p-değeri: 0.00921
H0 reddedilir: Müşteri memnuniyet oranı %90'dan düşüktür.
```

In [2]:

```python
# elimde veri olsaydı

data = [1]*170 + [0]*30  # 200 müşteriden 170 memnun (1), 30 memnun değil (0)

from statsmodels.stats.proportion import proportions_ztest

# Toplam memnun olan sayısı ve örneklem büyüklüğü
count = sum(data)         # 170
nobs = len(data)          # 200

# H0: p = 0.90
test_stat, p_value = proportions_ztest(count, nobs, value=0.90, alternative='sma

print(f"Z istatistiği: {test_stat:.3f}")
print(f"p-değeri: {p_value:.5f}")

if p_value < 0.05:
    print("H0 reddedilir: Müşteri memnuniyet oranı %90'dan düşüktür.")
else:
    print("H0 reddedilemez.")
```

```
Z istatistiği: -1.980
p-değeri: 0.02384
H0 reddedilir: Müşteri memnuniyet oranı %90'dan düşüktür.
```

📊 Neden Veri Varken ve Yokken Z Değerleri Farklı Çıkar?

Z-Testi'nde kullanılan **standart hata** (standard error) bileşeni, Z skorunu doğrudan etkiler. Bu bileşen, elimizdeki bilginin türüne göre farklı hesaplanır:

🎯 Temel Sebep: Standart Sapma'nın Hesaplanma Yöntemi

Durum	Kullanılan Standart Sapma	Etki
✅ Veri yokken	Hipotez oranı (p_0)'a göre	Daha küçük varyans → daha büyük Z
✅ Veri varken	Örneklem oranı (\hat{p})'ye göre	Daha büyük varyans → daha küçük Z

🔢 Sayısal Örnek

- Toplam müşteri sayısı: **n = 200**
- Memnun olan müşteri sayısı: **x = 170**
- Hipotezdeki memnuniyet oranı: **p_0 = 0.90**
- Örneklem oranı: (\hat{p} = 170 / 200 = 0.85)

✴️ 1. El ile (p_0'ye göre) hesaplanan Z

$$z = \frac{0.85 - 0.90}{\sqrt{\frac{0.90(1-0.90)}{200}}} = \frac{-0.05}{\sqrt{0.00045}} \approx -2.357$$

✴️ 2. Veri ile ($\hat{p}$'ye göre) hesaplanan Z

$$z = \frac{0.85 - 0.90}{\sqrt{\frac{0.85(1-0.85)}{200}}} = \frac{-0.05}{\sqrt{0.0006375}} \approx -1.980$$

📌 Neden Farklı?

- İlk hesaplama (veri **yokken**) sabit varsayıma dayanır → standart hata **daha küçük**, z **daha büyük** çıkar.
- İkinci hesaplama (veri **varken**) örneklem varyansına dayanır → daha fazla belirsizlik, **daha küçük z skoru** üretir.

🧠 Özet

Test Tipi	Varyans Kaynağı	Z Skoru	Kullanım Durumu
Teorik / Elle	(p_0 (1 - p_0)/n)	Daha büyük	Veri yoksa, teoriye dayalı test
Veri ile (`ztest` , `proportions_ztest`)	$(\hat{p}(1-\hat{p})/n)$	Daha küçük	Veri varsa, gerçek uygulamalarda

---

✅ Not:

Her iki yaklaşım da doğrudur.
Ancak **veri varken** yapılan hesaplama genellikle **daha güvenli ve temkinli** sonuç verir, çünkü örneklem varyansını hesaba katar.

T-Testi: Popülasyon Standart Sapması Bilinmediğinde

Şimdi sıra pratikte en sık kullanılan t-testinde.

T-Testine Neden İhtiyaç Duyarız? Z-Testinde Ne Yanlış Var?

Sorun şu: Gerçek hayatta σ'yi (popülasyon standart sapmasını) neredeyse hiç bilemeyiz. İşte t-testi tam da bu noktada devreye girer.

Sorun: Z testi formülünde σ yerine s (örneklem standart sapması) kullanıldığında, ek belirsizlik yaratılmış olur. Sadece ortalamadan değil, değişkenlikten de emin olunamaz.

Çözüm: William Gosset (Guinness bira fabrikasında çalışmış ve "Öğrenci" takma adıyla yayınlar yapmıştır) adlı bir adam bunu 1908'de buldu. Bu ekstra belirsizliği hesaba katan t-dağılımını geliştirdi.

Z ve T Arasındaki Temel Farklar:

Bakış Açısı	Z-Test	T-Test
Popülasyon Standart Sapması	Bilinen (σ)	Bilinmeyen(s kullanın)
Gerekli örneklem büyüklüğü	Büyük ($n \geq 30$) genellikle	Her boyutta çalışır
Dağılım Şekli	Standart normal (her zaman aynı)	Örneklem büyüklüğüne bağlı değişiklikler
Ne zaman kullanılır?	Pratikte nadirdir	Çok yaygın

T-Dağılımı

t-dağılımının görünümü normal dağılıma (çan eğrisi) benzer, ancak daha kalın kuyruklara sahiptir. Bu, daha temkinli olduğu anlamına gelir; daha fazla belirsizlik olduğunu kabul eder.

Görsel benzetme:

- Normal dağılım: Kendine güvenen bir okçunun atış düzeni - sıkı gruplama
- t-dağılımı (küçük örneklem): Daha az güvenen bir okçu - daha geniş yayılma

Örneklem büyüklüğü büyüdükçe, t-dağılım normal dağılıma yaklaşır. $n = 30$ civarında, neredeyse aynı olurlar. İşte bu yüzden insanlar "büyük örneklemler için z-testi, küçük örneklemler için t-testi kullanın" derler - ama aslında t-testi her ikisi için de işe yarar!

T-Test Türleri

Üç ana t-testi türü vardır ve bu, kafa karışıklığının en çok yaşandığı yerdir.

1. One-Sample T-Test

Cevapladığı soru: "Örneklem ortalaması varsayılan değerden anlamlı derecede farklı mı?"

Ne zaman kullanılır: Bir grubunuz olduğunda ve bunu belirli bir değerle karşılaştırmak istediğinizde.

Formül

$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$\bar{x}$: Örneklem Ortalaması μ_0 : Varsayılan ortalama s = örnek standart sapması n = örneklem büyüklüğü

Dikkat edin, bu z-testi formülüne çok benziyor, ancak σ yerine s kullanılıyor!

Örnek:

Senaryo: Bir gıda şirketi, mısır gevreği kutularının 500 gr mısır gevreği içerdiğini iddia ediyor. Siz de yetersiz doldurduklarından şüpheleniyorsunuz. Rastgele seçtiğiniz 25 kutuyu tartıyorsunuz.

- Data:
  - Örneklem büyüklüğü: n = 25 kutu
  - Örneklem Ortalaması $\bar{x}$ = 495 gr
  - Örnek standart sapması: s = 12g
- Adım 1: Hipotezler 1. H_0: μ ≥ 500g (Kutular uygun şekilde doldurulmuştur) 2. H_a: μ < 500g (Kutular yetersiz doldurulmuştur) Bu sol kuyruklu bir testtir.
- t-istatistiğini hesapla t = (495 - 500) / (12 / √25) t = -5 / (12 / 5) t = -5 / 2.4 t = -2.08
- Degrees of freedom df = n - 1 = 25 - 1 = 24
- Adım 4: Karar df = 24 için bir t-tablosuna bakıldığında (veya yazılım kullanıldığında), α = 0,05'te sol kuyruklu test: Kritik değer = -1,711
- -2.08 < -1.711 (daha uç) olduğundan H_0'yi reddediyoruz.

Sonuç: "Kutuların yetersiz doldurulduğuna dair güçlü kanıtlar var. Gerçek ortalama ağırlık muhtemelen 500 gramdan azdır."

In [3]:
```python
from math import sqrt
from scipy.stats import t

# Verilen bilgiler
x_bar = 495      # örneklem ortalaması
mu0 = 500        # hipotez edilen popülasyon ortalaması
s = 12           # örneklem standart sapması
n = 25           # örneklem büyüklüğü

# t istatistiği
t_stat = (x_bar - mu0) / (s / sqrt(n))
```

```python
# Tek kuyruklu p-değeri (H1: ortalama < mu0)
df = n - 1  # serbestlik derecesi
p_value = t.cdf(t_stat, df)

print(f"t istatistiği: {t_stat:.3f}")
print(f"p-değeri (tek kuyruklu): {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: Ortalama ağırlık 500 gramdan daha azdır.")
else:
    print("H0 reddedilemez: Veriler 500 gramdan düşük olduğuna dair yeterli kanı
```

```
t istatistiği: -2.083
p-değeri (tek kuyruklu): 0.02402
H0 reddedilir: Ortalama ağırlık 500 gramdan daha azdır.
```

In [5]:

```python
import numpy as np
from scipy.stats import ttest_1samp

# Örnek veri: 25 kutunun ağırlığı (gram cinsinden)
weights = np.array([
    497, 493, 494, 496, 492, 498, 495, 491, 496, 497,
    494, 490, 493, 492, 495, 496, 493, 497, 489, 492,
    495, 493, 490, 491, 494
])

# Hipotez edilen ortalama
mu0 = 500

# Tek örneklem t-testi
t_stat, p_left_tailed = ttest_1samp(weights, popmean=mu0, alternative='less')

# Tek kuyruklu p-değeri (H1: ortalama < mu0)
from scipy.stats import t
n = len(weights)
df = n - 1



# Sonuçları yazdır
print(f"t istatistiği: {t_stat:.3f}")
print(f"p-değeri (tek kuyruklu, H1: ort < 500): {p_left_tailed:.5f}")

# Karar
alpha = 0.05
if p_value_one_tailed < alpha:
    print("H0 reddedilir: Ortalama ağırlık 500 gramdan düşüktür.")
else:
    print("H0 reddedilemez: Ortalama ağırlığın düşük olduğuna dair yeterli kanıt
```

```
t istatistiği: -12.686
p-değeri (tek kuyruklu, H1: ort < 500): 0.00000
H0 reddedilir: Ortalama ağırlık 500 gramdan düşüktür.
```

2. Two-Sample T-Test (Independent Samples)

İşte asıl karışıklık burada başlıyor!

Cevapladığı soru: "İki bağımsız grubun ortalamaları anlamlı derecede farklı mıdır?" **Ne zaman kullanılır:** İki ayrı, ilgisiz grubunuz olduğunda ve bunların ortalamalarını karşılaştırmak istediğinizde.

Anahtar kelime: BAĞIMSIZ

- A Grubu insanları B Grubu insanlarından tamamen farklıdır
- Bir gruba olan şey diğerini etkilemez
- Gözlemler arasında doğal bir eşleşme yok

Formül (eşit varyanslar varsayılarak): $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$ -> (birleştirilmiş varyans)pooled variance

Burada neler oluyor:

- $(\bar{x}_1 - \bar{x}_2)$: İki örnek ortalaması arasındaki fark
- s_p^2 : Daha iyi bir tahmin elde edebilmek için her iki gruptan elde edilen varyanslar birleştirilir.
- Payda: Ortalamalar arasındaki farkın standart hatası

Yaygın Karışıklık #1: "Farklılıkları neden birleştiriyoruz?"

Diyelim ki 10 erkek ve 10 kadının boyları karşılaştırılıyor. Her grup size boylardaki değişkenlik hakkında bilgi verir. Sadece bir grubun değişkenliğini kullanmak yerine, daha istikrarlı bir tahmin elde etmek için her iki grubun bilgileri birleştirilir. Bu da daha fazla istatistiksel güç sağlar.

Not: Eşit varyansları varsaymayan (Welch t-testi olarak adlandırılır) ve pratikte kullanımı daha güvenli olan bir versiyonu da mevcuttur. Çoğu yazılım bunu otomatik olarak yapar.

Örnek : Yeni bir öğretim yönteminin (Yöntem A) geleneksel yöntemden (Yöntem B) daha iyi olup olmadığını öğrenmek istiyorsunuz. 30 öğrenciyi rastgele Yöntem A'ya, 28 öğrenciyi ise Yöntem B'ye atıyorsunuz. Bir dönem sonra, sınav puanlarını karşılaştırıyorsunuz.

Veriler

- Yöntem A : n_1 = 30, $(\bar{x}_1$ = 78, s_1 = 8
- Yöntem B : n_2 = 28, $(\bar{x}_2$ = 72, s_2 = 10

- **Bağımsızlığı doğrulayın ✓** A Yöntemindeki öğrenciler, B Yöntemindeki öğrencilerden farklı kişilerdir ✓ Rastgele atanırlar ✓ Bir öğrencinin puanı diğerinin puanını etkilemez. Bu, bağımsız iki örneklemli bir durumdur.
- Hipotezler: H_0: $\mu_1 = \mu_2$ (Yöntemler arasında fark yok) H_a: $\mu_1 \neq \mu_2$ (Yöntemler farklı sonuçlar üretir)

İki taraflı test (herhangi bir fark olup olmadığını kontrol etmek için).

Birleştirilmiş varyansı hesapla: $s\_p^2 = ((30-1)\times 8^2 + (28-1)\times 10^2) / (30 + 28 - 2)$ $s\_p^2 = (29\times 64 + 27\times 100) / 56$ $s\_p^2 = (1{,}856 + 2{,}700) / 56$ $s\_p^2 = 4{,}556 / 56$ $s\_p^2 = 81.36$

t-istatistiğini hesapla: $t = (78 - 72) / \sqrt{(81.36 \times (1/30 + 1/28))}$ $t = 6 / \sqrt{(81.36 \times 0.0690)}$ $t = 6 / \sqrt{5.61}$ $t = 6 / 2.37$ $t = 2.53$

Degrees of freedom $df = n_1 + n_2 - 2 = 30 + 28 - 2 = 56$

$df = 56$ için, $\alpha = 0{,}05$'te iki kuyruklu test: Kritik değerler $= \pm 2{,}003$

$2{,}53 > 2{,}003$ olduğundan H_0'yi reddediyoruz

Sonuç: "Yöntem A ve Yöntem B'nin farklı ortalama puanlar ürettiğine dair önemli kanıtlar mevcuttur. Yöntem A'nın daha etkili olduğu (daha yüksek ortalama puan) görülmektedir."

In [7]:
```
from math import sqrt
from scipy.stats import t

# Veriler
x1, s1, n1 = 78, 8, 30
x2, s2, n2 = 72, 10, 28

# Pooled variance hesapla (eş varyans varsayımı)
sp_squared = ((n1 - 1)*s1**2 + (n2 - 1)*s2**2) / (n1 + n2 - 2)

# t istatistiği hesapla
t_stat = (x1 - x2) / sqrt(sp_squared * (1/n1 + 1/n2))

# Serbestlik derecesi
df = n1 + n2 - 2

# Çift kuyruklu p-değeri
p_value = 2 * (1 - t.cdf(abs(t_stat), df))

# Sonuçları yazdır
print(f"t istatistiği: {t_stat:.3f}")
print(f"p-değeri (çift kuyruklu): {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: Yöntemler arasında anlamlı bir fark vardır.")
else:
    print("H0 reddedilemez: Yöntemler arasında anlamlı fark yoktur.")
```

```
t istatistiği: 2.532
p-değeri (çift kuyruklu): 0.01419
H0 reddedilir: Yöntemler arasında anlamlı bir fark vardır.
```

In [8]:
```
import numpy as np
from scipy.stats import ttest_ind

# Yapay veri üretimi (aynı ortalama ve std'lerle)
np.random.seed(0)
group_A = np.random.normal(loc=78, scale=8, size=30)
group_B = np.random.normal(loc=72, scale=10, size=28)

# İki örneklem t-testi (eş varyanslı, çift kuyruklu)
t_stat, p_value = ttest_ind(group_A, group_B, equal_var=True)

# Sonuçları yazdır
print(f"t istatistiği: {t_stat:.3f}")
```

```python
print(f"p-değeri (çift kuyruklu): {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: Yöntemler arasında anlamlı bir fark vardır.")
else:
    print("H0 reddedilemez: Yöntemler arasında anlamlı fark yoktur.")
```

```
t istatistiği: 5.128
p-değeri (çift kuyruklu): 0.00000
H0 reddedilir: Yöntemler arasında anlamlı bir fark vardır.
```

Paired T-Test (Dependent Samples)

Bu, t-testlerindeki EN çok karıştırılan konudur!

Cevapladığı soru: "İki ilişkili ölçüm arasında anlamlı bir fark var mıdır?"

Ne zaman kullanılır: Aynı denek üzerinde iki ölçümünüz olduğunda veya doğal olarak eşleştirilmiş gözlemleriniz olduğunda.

Anahtar kelime: EŞLEŞMİŞ/BAĞIMLI

- Aynı kişilere iki kez ölçüm yapıldı (önce/sonra)
- Eşleştirilmiş çiftler (ikizler, karı-koca, sol el-sağ el)
- Doğal eşleşme bir bağımlılık yaratır

Kritik İçgörü

Eşleştirilmiş t-testinde, iki ayrı grubu KARŞILAŞTIRMAZSINIZ. Her bir çift içindeki farka bakarsınız ve ardından ortalama farkın sıfırdan anlamlı derecede farklı olup olmadığını test edersiniz.

Formül $t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$

$\bar{d}$: farkların ortalaması s_d = farkların standart sapması n = çift sayısı

Bu, tıpkı tek örneklemli t-testi gibi görünüyor! Çünkü aslında farklılıklar üzerinde tek örneklemli bir t-testi.

Yaygın Karışıklık #2: "Eşleştirilmiş ve bağımsız t-testini ne zaman kullanmalıyım?"

Kendinize şunu sorun: "Doğal çiftler var mı?"

EŞLEŞTİRİLMİŞ durumların örnekleri:

1. Öncesi/Sonrası: Aynı kişilerin diyet öncesi ve sonrası kiloları
2. Sol ve Sağ: Aynı kişilerin sol ve sağ ellerinin gücü
3. Eşleştirilmiş Çiftler: Aynı hanede kocanın geliri ile karısının geliri
4. Aynı Konu, Farklı Koşullar: Aynı kişinin kafeinli ve kafeinsiz tepki süresi

BAĞIMSIZ durumların örnekleri:

1. Erkeklerin boyları ve kadınların boyları (farklı kişiler)

2. A Sınıfı ve B Sınıfı öğrencilerinin sınav puanları (farklı öğrenciler)
3. Mağaza 1 ve Mağaza 2'deki satışlar (farklı mağazalar)

Anahtar Soru: "Grup 2'nin sırasını karıştırırsam, fark eder mi?"

- Eşleştirilmiş: EVET! 1. Gruptaki her gözlemin 2. Grupta belirli bir ortağı vardır.
- Bağımsız: HAYIR! 1. Gruptaki herhangi bir kişi, 2. Gruptaki herhangi bir kişiyle karşılaştırılabilir.

Örnek

Senaryo: Bir tansiyon ilacının işe yarayıp yaramadığını test etmek istiyorsunuz. İlacı almadan ÖNCE 20 hastanın tansiyonunu ölçüyorsunuz, ardından tedaviden bir ay SONRA tekrar ölçüyorsunuz.

Bu neden EŞLEŞTİRİLDİ:

- Her hasta kendi kontrolü olarak hizmet eder
- Hasta 1'in "önce" ölçümü, doğal olarak Hasta 1'in "sonra" ölçümüyle eşleştirilir
- Her hastanın İÇİNDEKİ değişimle ilgileniyoruz

Veri: verilerin bir alt kümesi:

Hasta	Önce	Sonra	Fark(Önce-Sonra)
1	145	138	7
2	152	145	7
3	138	135	3
4	160	150	10
5	142	140	2
...	...	...	...

20 hastanın tamamı için:

- Ortalama fark: $\bar{d}$ = 6.5 mmHg
- Farkların standart sapması: s_d = 4.2
- n = 20 hasta

Adım 1: Hipotezler

- H_0: μ_d = 0 (İlaç kan basıncını değiştirmez)
- H_a: μ_d > 0 (İlaç kan basıncını düşürür)

Not: Ortalama farkın sıfırdan büyük olup olmadığını (sağ kuyruklu) test ediyoruz.

Adım 2: t-istatistiğini hesaplayın t = (6.5 - 0) / (4.2 / √20) t = 6.5 / (4.2 / 4.47) t = 6.5 / 0.94 t = 6.91

Adım 3: Serbestlik derecesi df = n - 1 = 20 - 1 = 19

Adım 4: Karar df = 19 için, α = 0,05'te sağ kuyruklu test: Kritik değer = 1,729 6.91 > 1.729 olduğundan (çok daha uç!), H_0'yi kesinlikle reddediyoruz.

Sonuç: "İlacın kan basıncını düşürdüğüne dair çok güçlü kanıtlar var. Hastalarda ortalama 6,5 mmHg'lik bir düşüş gözlemlendi."

Eşleştirilmiş analiz neden önemlidir? Eşleştirmeyi göz ardı edip bağımsız iki örneklemli t-testi kullansaydık, bilgiyi çöpe atmış olurduk! Eşleştirilmiş testin gücü, bireysel çeşitliliği kontrol etmesinden gelir. Bazı insanların kan basıncı doğal olarak diğerlerinden daha yüksektir, ancak biz bunu önemsemiyoruz; her bir kişinin içindeki DEĞİŞİMLE ilgileniyoruz.

Yaygın Karışıklık #3: "İki grubumun boyutları farklıysa eşleştirilmiş bir test kullanabilir miyim?"

HAYIR! Tanımı gereği bu imkansızdır.

A Grubu'nda 30, B Grubu'nda 25 kişi varsa, onları eşleştirmenin bir yolu yoktur! Bağımsız iki örneklemli t-testi KULLANMALISINIZ.

Çiftli testler her grupta tam olarak aynı sayıda gözlem gerektirir çünkü her gözlemin bir partnere ihtiyacı vardır.

Yaygın Karışıklık #4: "Verilerim eşleştirilmiş, ancak yine de normalliği kontrol etmeli miyim?" EVET! Eşleştirilmiş t-testi için, farkların yaklaşık olarak normal dağılım göstermesi gerekir (veya yeterince büyük bir örneğiniz olmalıdır).

"Önce" ve "sonra" değerlerinin kendilerinin normal olup olmaması sizin için önemli değil; sadece farkların normal olup olmaması önemli.

In [2]:
```
from math import sqrt
from scipy.stats import t

# Verilen bilgiler
d_bar = 6.5   # farkların ortalaması
sd = 4.2      # farkların standart sapması
n = 20        # hasta sayısı

# t istatistiği
t_stat = (d_bar - 0) / (sd / sqrt(n))

# Serbestlik derecesi
df = n - 1

# Tek kuyruklu p-değeri (H1: d > 0)
p_value = 1 - t.cdf(t_stat, df)

# Sonuçları yazdır
print(f"t istatistiği: {t_stat:.3f}")
print(f"p-değeri (tek kuyruklu): {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: İlaç kan basıncını anlamlı derecede düşürmektedir.")
```

```
else:
    print("H0 reddedilemez: İlacın tansiyonu düşürdüğüne dair yeterli kanıt yokt
```

```
t istatistiği: 6.921
p-değeri (tek kuyruklu): 0.00000
H0 reddedilir: İlaç kan basıncını anlamlı derecede düşürmektedir.
```

In [1]:

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

# Örnek veri oluşturma (simülasyon)
# "Önce" tansiyon ortalaması yaklaşık 140, "Sonra" ortalaması yaklaşık 133.5
np.random.seed(42)
before = np.random.normal(loc=140, scale=8, size=20)
after = before - np.random.normal(loc=6.5, scale=4.2, size=20)

# Paired t-test
t_stat, p_two_tailed = ttest_rel(before, after)

# Tek kuyruklu p-değeri (H1: before > after yani ilaç -> düşürür)
if t_stat > 0:
    p_one_tailed = p_two_tailed / 2
else:
    p_one_tailed = 1 - (p_two_tailed / 2)

# Sonuçları yazdır
print(f"Önce ortalama: {np.mean(before):.2f}")
print(f"Sonra ortalama: {np.mean(after):.2f}")
print(f"t istatistiği: {t_stat:.3f}")
print(f"p-değeri (tek kuyruklu): {p_one_tailed:.7f}")

# Karar
alpha = 0.05
if p_one_tailed < alpha:
    print("H0 reddedilir: İlaç kan basıncını anlamlı derecede düşürmektedir.")
else:
    print("H0 reddedilemez: İlacın tansiyonu düşürdüğüne dair yeterli kanıt yokt
```

```
The history saving thread hit an unexpected error (OperationalError('attempt to w
rite a readonly database')).History will not be written to the database.
Önce ortalama: 138.63
Sonra ortalama: 133.25
t istatistiği: 5.921
p-değeri (tek kuyruklu): 0.0000053
H0 reddedilir: İlaç kan basıncını anlamlı derecede düşürmektedir.
```

Ki-Kare (χ^2) Testleri: Kategorik Veriler İçin

Şimdi tamamen farklı bir veri türüne geçiyoruz: sayılar yerine kategoriler.

Ki-Kare Testlerini Ne Zaman Kullanırız?

Şimdiye kadar ortalamaları karşılaştırdık. Peki ya verileriniz sayısal değilse? Ya farklı kategorilere kaç kişinin girdiğini sayıyorsanız?

Örnekler

- Kırmızı arabalar, mavi arabalar, beyaz arabalar

- Ürün A'yı ve Ürün B'yi Ürün C'ye tercih eden kişiler
- Hastalar ve sağlıklılar
- Başarılılar ve başarısızlar

Ki-kare testleri bunlar için mükemmeldir!

Üç Tür Ki-Kare Testi

Bu durum insanları şaşırtıyor çünkü testlerin isimleri benzer ama cevapları farklı. Her birini açıklayalım.

1. Chi-Square Goodness of Fit Test (Ki-Kare Uyum İyiliği Testi)

Cevapladığı soru: "Gözlemlediğim kategorilerin dağılımı beklenen dağılımla uyuşuyor mu?"

Ne zaman kullanılır: BİR kategorik değişkeniniz olduğunda ve frekansların varsayılan bir desenle eşleşip eşleşmediğini test etmek istediğinizde.

Basit Bir Benzetme: Bir zarı 60 kez atıyorsunuz. Zar adilse, her sayının (1-6) yaklaşık 10 kez gelmesini beklersiniz. Ama aslında şunu elde edersiniz:

- 1: 8 kere
- 2: 12 kere
- 3: 9 kere
- 4: 11 kere
- 5: 13 kere
- 6: 7 kere

"Her biri 10"dan sapmalar sadece rastgele bir değişim mi, yoksa zarın bir etkisi mi var?

Formül:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O = Gözlenen frekans E = Beklenen frekans Σ = Tüm kategorilerdeki toplam

Bu formülün anlamı:

- Her kategori için şunları hesaplayın: (Elde ettiğiniz sonuç - Beklediğiniz sonuç)² / Beklediğiniz sonuç
- Tüm bu değerleri toplayın
- Büyük χ^2, gözlemlenen verilerinizin beklenenden çok farklı olduğu anlamına gelir
- Küçük χ^2, gözlemlenen verilerinizin beklenene yakın olduğu anlamına gelir

Örnek Senaryo: Bir şirket M&M şekerlemeleri üretiyor. Dağıtımın şu şekilde olduğunu iddia ediyorlar:

- %30 Kahverengi
- %20 Sarı
- %20 Kırmızı

- %10 Turuncu
- %10 Yeşil
- %10 Mavi

İçinde 500 adet M&M bulunan büyük bir çanta satın alıyorsunuz ve iddialarının doğru olup olmadığını test etmek istiyorsunuz.

Gözlemlediğiniz sayılar:

Renk	Gözlemlenen(O)	Beklenen(E)
Kahverengi	140	150 (30% of 500)
Sarı	110	100 (20% of 500)
Kırmızı	95	100 (20% of 500)
Turuncu	60	50 (10% of 500)
Yeşil	50	50 (10% of 500)
Mavi	45	50 (10% of 500)
Toplam	500	500

Adım 1: Hipotezler

● H_0: Dağılım iddia edilen oranlarla eşleşiyor ● H_a: Dağılım iddia edilen oranlarla eşleşmiyor

Adım 2: χ^2'yi hesaplayın

Kahverengi: $(140 - 150)^2 / 150 = 100 / 150 = 0.67$ Sarı: $(110 - 100)^2 / 100 = 100 / 100 = 1.00$ Kırmızı: $(95 - 100)^2 / 100 = 25 / 100 = 0.25$ Turuncu: $(60 - 50)^2 / 50 = 100 / 50 = 2.00$ Yeşil: $(50 - 50)^2 / 50 = 0 / 50 = 0.00$ Mavi: $(45 - 50)^2 / 50 = 25 / 50 = 0.50$

$\chi^2 = 0.67 + 1.00 + 0.25 + 2.00 + 0.00 + 0.50 = 4.42$

Adım 3: Serbestlik derecesi df = (kategori sayısı) - 1 = 6 - 1 = 5

Adım 4: Karar df = 5 ve $\alpha = 0{,}05$ için, χ^2 tablosundan kritik değer = 11,07

$4.42 < 11.07$ olduğundan H_0'yi reddedemiyoruz.

Sonuç: "Gözlemlenen dağılım, şirketin iddia ettiği oranlarla tutarlıdır. Herhangi bir sapma, rastgele şansa bağlanabilir."

In [4]:
```
import numpy as np
from scipy.stats import chi2

# Gözlenen ve beklenen değerler
observed = np.array([140, 110, 95, 60, 50, 45])
expected = np.array([150, 100, 100, 50, 50, 50])

# Ki-kare istatistiği manuel hesaplama
chi_square = np.sum((observed - expected)**2 / expected)
```

```python
# Serbestlik derecesi = kategori sayısı - 1
df = len(observed) - 1

# p-değeri
p_value = 1 - chi2.cdf(chi_square, df)

print(f"Chi-Square istatistiği: {chi_square:.3f}")
print(f"Serbestlik derecesi: {df}")
print(f"p-değeri: {p_value:.5f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("H0 reddedilir: Dağılım iddia edilenden farklıdır.")
else:
    print("H0 reddedilemez: Dağılım iddia edilen dağılımla uyumludur.")
```

```
Chi-Square istatistiği: 4.417
Serbestlik derecesi: 5
p-değeri: 0.49111
H0 reddedilemez: Dağılım iddia edilen dağılımla uyumludur.
```

In [5]:

```python
from scipy.stats import chisquare

chi_stat, p_value = chisquare(f_obs=observed, f_exp=expected)

print(f"Chi-Square istatistiği: {chi_stat:.3f}")
print(f"p-değeri: {p_value:.5f}")

if p_value < 0.05:
    print("H0 reddedilir: Dağılım iddia edilenden farklıdır.")
else:
    print("H0 reddedilemez: Dağılım iddia edilen dağılımla uyumludur.")
```

```
Chi-Square istatistiği: 4.417
p-değeri: 0.49111
H0 reddedilemez: Dağılım iddia edilen dağılımla uyumludur.
```

Burada hangi kuyruk testinin yapıldığının önemi var mı? ✅

🎯 Cevap: **Hayır, hangi kuyruk olduğunun önemi yok**

Çünkü **Ki-Kare testi daima tek yönlüdür (right-tailed / sağ kuyruk)**.

🧠 Neden?

Ki-Kare istatistiği şu şekilde hesaplanır:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Bu formül:

- Daima **pozitif** bir değer üretir
- Gözlenen değerler **beklenenden ne kadar uzaksa** ki-kare büyür

- Dağılım **negatif değer almaz**

Dolayısıyla:

> Sadece **"çok büyük farklar var mı?"** sorusu test edilir.

Yani:

- Büyük fark → **χ^2 büyük olur** → **H_0 reddedilir**
- Küçük fark → **χ^2 küçük olur** → **H_0 reddedilemez**

Bu nedenle ki-kare dağılımı:

✅ **Tek kuyruklu bir dağılımdır** ❌ Sol kuyruk hiç kullanılmaz

---

📈 Görsel olarak

Yüksek χ^2 → sağ kuyruğa düşer → p küçük → *fark var* Düşük χ^2 → merkezde → p büyük → *fark yok*

🔎 Özet Tablo

Test	Kuyruk Türü
Z-Test	Tek veya çift kuyruk olabilir
t-Test	Tek veya çift kuyruk olabilir
χ^2 Goodness of Fit Test	✅ Her zaman sağ kuyruk
χ^2 Independence Test	✅ Her zaman sağ kuyruk

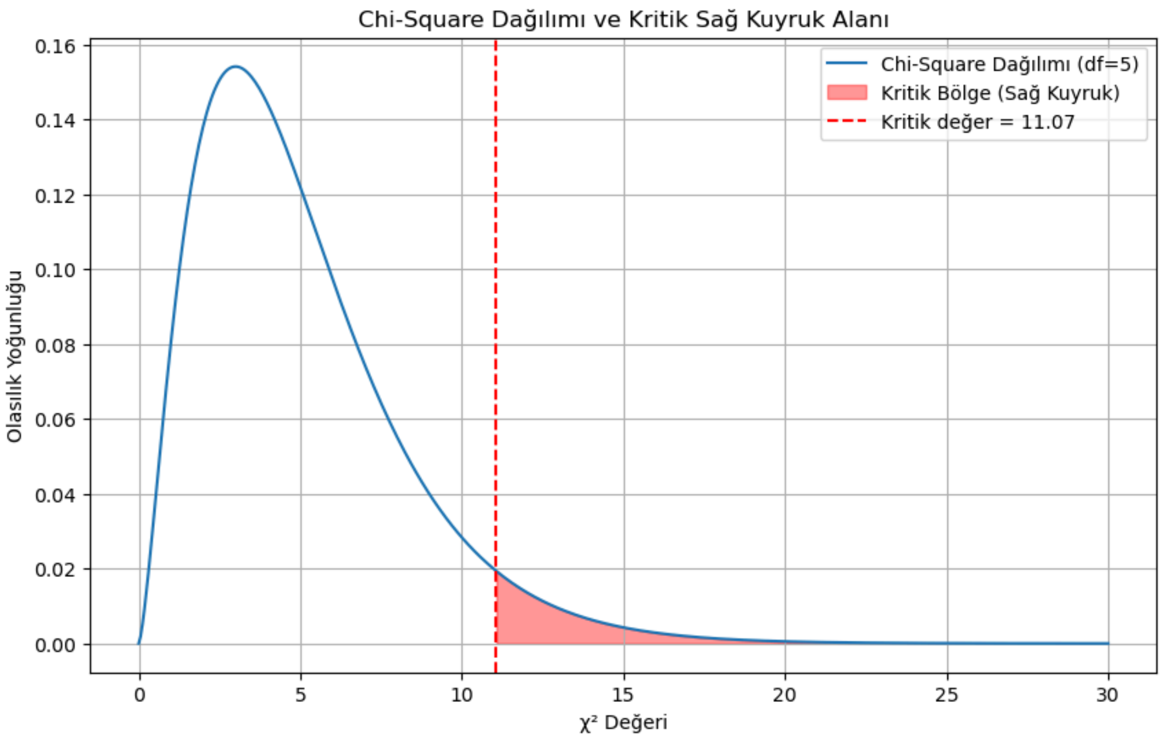
✅ Sonuç

> **Ki-Kare testlerinde "hangi kuyruk?" diye düşünmeye gerek yok. Ki-Kare = her zaman sağ kuyruklu testtir.**

In [6]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import chi2

# Serbestlik derecesi
df = 5
alpha = 0.05

# Kritik değer (sağ kuyruk sınırı)
critical_value = chi2.ppf(1 - alpha, df)

# x değerleri
```

```python
x = np.linspace(0, 30, 500)

# Ki-Kare pdf (olasılık yoğunluğu)
pdf = chi2.pdf(x, df)

plt.figure(figsize=(10, 6))
plt.plot(x, pdf, label=f'Chi-Square Dağılımı (df={df})')

# Sağ kuyruğu renklendir
x_filled = np.linspace(critical_value, 30, 200)
plt.fill_between(x_filled, chi2.pdf(x_filled, df), color='red', alpha=0.4,
                 label='Kritik Bölge (Sağ Kuyruk)')

# Kritik değeri işaretle
plt.axvline(critical_value, color='red', linestyle='--',
            label=f'Kritik değer = {critical_value:.2f}')

plt.title('Chi-Square Dağılımı ve Kritik Sağ Kuyruk Alanı')
plt.xlabel('χ² Değeri')
plt.ylabel('Olasılık Yoğunluğu')
plt.legend()
plt.grid(True)
plt.show()
```



2. Chi-Square Test of Independence

Cevapladığı soru: "İki kategorik değişken birbirleriyle ilişkili midir?"

Ne zaman kullanılır: İKİ kategorik değişkeniniz olduğunda ve bunların ilişkili olup olmadığını bilmek istediğinizde.

Basit Bir Benzetme: İnsanlara şu konuda anket yapıyorsunuz:

- Değişken 1: Cinsiyet (Erkek/Kadın)
- Değişken 2: Tercih edilen içecek (Kahve/Çay)

Cinsiyet ve içecek tercihi ilişkili midir? Yoksa bağımsız mıdırlar (ilişkili değillerdir)?

Yaygın Karışıklık #5: "Bu, Uyum İyiliğinden nasıl farklıdır?"

Goodness of Fit: BİR değişken, beklenen oranlara karşı test. Örnek: "İnsanlar kırmızıyı, maviyi veya yeşili eşit olarak mı tercih ediyor?" **Test of Independence:** İKİ değişken, bunların ilişkili olup olmadığının test edilmesi. Örnek: "Renk tercihi cinsiyetle ilişkili midir?"

Veri Formatı: Bağımsızlık testleri için verileri bir olasılık tablosunda (çapraz tablolama olarak da adlandırılır) düzenlenir.

Örnek

Senaryo: Sigara içme durumu ile akciğer hastalığı arasında bir ilişki olup olmadığını öğrenmek istiyorsunuz. 500 kişiye anket uyguluyorsunuz.

Gözlemlenen Veriler (Olasılık Tablosu):

	Akciğer Hastalığı Var	Akciğer Hastalığı Yok	Satır Toplamı
Sigara içen	60	140	200
Sigara İçmeyen	30	270	300
Sütun Toplamı	90	410	500

Adım 1: Hipotezler ● H_0: Sigara içme durumu ve akciğer hastalığı bağımsızdır (ilişkili değildir) ● H_a: Sigara içme durumu ve akciğer hastalığı bağımlıdır (ilişkili)

Adım 2: Beklenen Frekansları Hesaplayın İşte işin zor kısmı bu. İki değişken gerçekten bağımsız olsaydı, her hücrede ne kadar sayım beklerdik?

Beklenen frekans formülü: E = (Satır Toplamı × Sütun Toplamı) / Genel Toplam

"Sigara İçen + Akciğer Hastalığı" hücresi için: E = (200 × 90) / 500 = 18.000 / 500 = 36 "Sigara İçen + Akciğer Hastalığı Yok" hücresi için: E = (200 × 410) / 500 = 82.000 / 500 = 164 "Sigara İçmeyen + Akciğer Hastalığı" hücresi için: E = (300 × 90) / 500 = 27.000 / 500 = 54 "Sigara İçmeyen + Akciğer Hastalığı Yok" hücresi için: E = (300 × 410) / 500 = 123.000 / 500 = 246

Beklenen Frekanslar Tablosu:

	Akciğer Hastalığı Var	Akciğer Hastalığı Yok
Sigara içen	36	164
Sigara İçmeyen	54	246

Step 3: Calculate χ^2

Tüm hücreler için $\chi^2 = \sum \frac{(O-E)^2}{E}$

Hücre 1 (Sigara İçen, Hastalık): $(60 - 36)^2 / 36 = 576 / 36 = 16{,}00$ Hücre 2 (Sigara İçen, Hastalık Yok): $(140 - 164)^2 / 164 = 576 / 164 = 3{,}51$ Hücre 3 (Sigara İçmeyen, Hastalık): $(30 - 54)^2 / 54 = 576 / 54 = 10{,}67$ Hücre 4 (Sigara İçmeyen, Hastalık Yok): $(270 - 246)^2 / 246 = 576 / 246 = 2{,}34$

$\chi^2 = 16.00 + 3.51 + 10.67 + 2.34 = 32.52$

İlginç bir gözlem: Pay $(O - E)^2$'nin tüm hücreler için aynı olduğunu (576) fark ettiniz mi? Bu, sapmaların birbiriyle ilişkili olması nedeniyle 2×2 tablolarda görülür.

Adım 4: Serbestlik dereceleri

Bağımsızlık testleri için:

df = (satır sayısı - 1) × (sütun sayısı - 1) df = (2 - 1) × (2 - 1) = 1 × 1 = 1

Adım 5: Karar

df = 1 ve $\alpha = 0{,}05$ için kritik değer = 3,84 32,52 > 3,84 olduğundan (çok daha uç!), H_0'yi kesinlikle reddediyoruz.

Sonuç: "Sigara içme durumu ile akciğer hastalıkları arasında çok güçlü bir ilişki olduğuna dair kanıtlar var. Sigara içenlerin akciğer hastalığı oranı, değişkenler bağımsız olsaydı beklenenden çok daha yüksektir."

Pratik Yorumlama: Verilere bakıldığında: ● Sigara içenler arasında: 60/200 = %30'unda akciğer hastalığı var ● Sigara içmeyenler arasında: 30/300 = %10'unda akciğer hastalığı var

Bu örnekte sigara içenlerin akciğer hastalığına yakalanma olasılığı 3 kat daha fazla!

In [1]:
```python
import numpy as np
from scipy.stats import chi2_contingency

# Contingency table (gözlemlenen frekanslar)
observed = np.array([[60, 140],
                     [30, 270]])

# Chi-Square Test
chi2_stat, p_value, dof, expected = chi2_contingency(observed)

print("Chi-Square istatistiği:", round(chi2_stat, 3))
print("Serbestlik derecesi:", dof)
print("p-değeri:", round(p_value, 5))

# Beklenen frekans tablo çıktısı
print("\nBeklenen Değerler:")
print(expected)

# Karar
alpha = 0.05
if p_value < alpha:
    print("\nH0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında il
else:
    print("\nH0 reddedilemez: İki değişken arasında anlamlı bir ilişki yoktur.")
```

```
Chi-Square istatistiği: 31.179
Serbestlik derecesi: 1
p-değeri: 0.0

Beklenen Değerler:
[[ 36. 164.]
 [ 54. 246.]]

H0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında ilişki vardır.
```

In [2]:

```python
import numpy as np
import pandas as pd
from scipy.stats import chi2_contingency

# Ham veri oluşturma
data = []

data += [["Smoker", "Disease"]] * 60
data += [["Smoker", "No Disease"]] * 140
data += [["Non-Smoker", "Disease"]] * 30
data += [["Non-Smoker", "No Disease"]] * 270

df = pd.DataFrame(data, columns=["Smoking", "LungDisease"])

# Kontenjans tablosu oluştur
contingency_table = pd.crosstab(df["Smoking"], df["LungDisease"])
print("Gözlenen Kontenjans Tablosu:")
print(contingency_table)

# Ki-Kare Bağımsızlık Testi
chi2_stat, p_value, dof, expected = chi2_contingency(contingency_table)

print("\nChi-Square istatistiği:", round(chi2_stat, 3))
print("Serbestlik derecesi:", dof)
print("p-değeri:", round(p_value, 7))

print("\nBeklenen Değerler:")
print(pd.DataFrame(expected,
                   index=contingency_table.index,
                   columns=contingency_table.columns))

# Karar
alpha = 0.05
if p_value < alpha:
    print("\nH0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında an
else:
    print("\nH0 reddedilemez: Sigara içme ile akciğer hastalığı arasında ilişki
```

```
Gözlenen Kontenjans Tablosu:
LungDisease  Disease  No Disease
Smoking
Non-Smoker        30         270
Smoker            60         140

Chi-Square istatistiği: 31.179
Serbestlik derecesi: 1
p-değeri: 0.0

Beklenen Değerler:
LungDisease  Disease  No Disease
Smoking
Non-Smoker      54.0       246.0
Smoker          36.0       164.0

H0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında anlamlı ilişki v
ardır.
```

In [3]:

```python
import numpy as np
import pandas as pd
from scipy.stats import chi2

# Gözlenen frekans tablosu
observed = np.array([[60, 140],
                     [30, 270]])

# Satır ve sütun toplamları
row_totals = observed.sum(axis=1)
col_totals = observed.sum(axis=0)
grand_total = observed.sum()

# Beklenen değerleri hesapla
expected = np.outer(row_totals, col_totals) / grand_total

# χ² = Σ (O - E)^2 / E
chi_square = np.sum((observed - expected)**2 / expected)

# Serbestlik derecesi = (satır-1)*(sütun-1)
df = (observed.shape[0] - 1) * (observed.shape[1] - 1)

# p-değeri
p_value = 1 - chi2.cdf(chi_square, df)

# Sonuçları yazdır
print("Gözlenen Tablo:")
print(pd.DataFrame(observed, index=["Smoker", "Non-Smoker"], columns=["Disease",

print("\nBeklenen Tablo:")
print(pd.DataFrame(np.round(expected, 2), index=["Smoker", "Non-Smoker"], column

print(f"\nChi-Square İstatistiği: {chi_square:.3f}")
print(f"Serbestlik Derecesi: {df}")
print(f"p-değeri: {p_value:.7f}")

# Karar
alpha = 0.05
if p_value < alpha:
    print("\nH0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında il
```

```
else:
    print("\nH0 reddedilemez: Sigara içme durumu ile akciğer hastalığı arasında
```

```
Gözlenen Tablo:
            Disease  No Disease
Smoker           60         140
Non-Smoker       30         270

Beklenen Tablo:
            Disease  No Disease
Smoker         36.0       164.0
Non-Smoker     54.0       246.0

Chi-Square İstatistiği: 32.520
Serbestlik Derecesi: 1
p-değeri: 0.0000000

H0 reddedilir: Sigara içme durumu ile akciğer hastalığı arasında ilişki vardır.
```

chi2_contingency kullanarak ve kullanmadan yapıldığında test istatistik değerleri neden farklı çıkıyor?

Gözlemlenen fark aslında **hesaplama biçiminden** kaynaklanıyor. Bu durum özellikle ki-kare bağımsızlık testinde çok sık karşımıza çıkar.

✅ Temel Fark: **Yates Continuity Correction**

`scipy.stats.chi2_contingency()` fonksiyonu **varsayılan olarak** şu düzeltmeyi uygular:

> 📌 **Yates continuity correction** (süreklilik düzeltmesi)

Bu düzeltme, özellikle:

- **2×2 kontenjans tablolarında**
- **gözlenen frekanslar** görece küçük olduğunda

→ testin **daha konservatif (temkinli)** olmasını sağlar.

🔬 Ne değişiyor?

Bizim manuel hesapladığımız formül:

[\chi^2 = \sum \frac{(O - E)^2}{E}]

Fakat `chi2_contingency()` düzeltme uyguladığında:

[\chi^2 = \sum \frac{(|O - E| - 0.5)^2}{E}]

📌 Dikkat edersen **mutlak farktan 0.5 çıkarılıyor** → Bu **istatistiği küçültür** → Böylece **p-değeri daha büyük çıkar** → H_0'ı reddetmek daha zorlaşır ✅

🔍 Kısa Tablo Özeti

Yöntem	Formül	Zorluk Derecesi	Sonuç Etkisi		
Manuel hesaplama	((O-E)^2/E)	Daha hızlı reddeder	χ^2 daha büyük		
`chi2_contingency()` varsayılan	((	O-E	-0.5)^2/E)	Daha konservatif	χ^2 daha küçük ✅

📌 Bu düzeltme ne zaman uygulanır?

Durum	Düzeltme önerisi
2×2 tablo	✅ Yates düzeltmesi
Daha büyük tablolar (ör. 3×3)	❌ Uygulanmaz
Beklenen frekanslar düşükse	✅ Faydalıdır

✅ `chi2_contingency` içinde bu düzeltmeyi kapatmak istersen:

```
chi2_stat, p, dof, exp = chi2_contingency(observed, correction=False)
```

➡️ Bu durumda **manuel hesapla aynı sonucu** alırsın ✔️

🎯 Sonuç

> 🔸 Kullandığın fonksiyon **varsayılan olarak** süreklilik düzeltmesi yapar 🔸 Manuel formülünde bu düzeltme yok 🔸 Bu nedenle sonuçlar **biraz farklı çıkar**, ama hangisi doğru?

✅ Hangisini kullanmalıyım?

Senaryo	Önerilen
2×2 tablo + küçük gözlem sayıları	✅ Yates düzeltmesi
Büyük tablo veya büyük n	❌ Düzeltme gereksiz

Eğer istersen:

✅ Aynı tablo için Yates düzeltmeli ve düzeltmesiz sonuçları **yan yana** gösteren bir Jupyter hücresi yazabilirim ✅ Sonuçları görselleştirebilirim (heatmap + mosaic plot) ✅ Etki büyüklüğü (Cramér's V) hesaplayabilirim

Hangisini istersin? 😊

3. Chi-Square Test of Homogeneity

Cevapladığı soru: "Farklı popülasyonlar kategoriler arasında aynı dağılıma sahip midir?"
Ne zaman kullanılır: Birden fazla popülasyondan örnekleriniz olduğunda ve kategori dağılımlarının aynı olup olmadığını bilmek istediğinizde.

Yaygın Karışıklık #6: "Bunun Test of Independence'den farkı nedir?" Matematiksel olarak aynı formülü kullanıyorlar. Ama kavramsal olarak:

Test of Independence:

- TEK bir popülasyondan TEK bir örneklem
- Her bireyde ölçülen iki değişken
- Soru: "Bu iki değişken birbiriyle ilişkili mi?"
- Örnek: 500 kişiye anket yapın, hem cinsiyeti hem de içecek tercihlerini kaydedin

Test of Homogeneity:

- FARKLI popülasyonlardan ÇOKLU örneklem
- Her bireyde ölçülen bir değişken
- Soru: "Bu popülasyonlar aynı dağılıma sahip mi?"
- Örnek: 200 Hintli ve 200 Çinliyle ayrı ayrı anket yapın, her gruptaki içecek tercihlerini kaydedin

Temel Fark:

Bağımsızlıkta, insanları örnekleyip iki özelliğe göre sınıflandırırsınız. Homojenlikte ise, farklı gruplardan örnek alıp aynı özelliği gruplar arasında karşılaştırırsınız.

Örnek

Senaryo: Bir ilaç şirketi bir ilacı üç ülkede test ediyor: ABD, Hindistan ve Brezilya. Yan etki profilinin ülkeler arasında aynı olup olmadığını öğrenmek istiyorlar.

İlacı şu kişilere veriyorlar:

- ABD'de 150 hasta
- Hindistan'da 150 hasta
- Brezilya'da 150 hasta

her hastanın hiçbir yan etki yaşamadığını, hafif yan etki yaşadığını veya şiddetli yan etki yaşadığını kaydedin.

Gözlemlenen Veriler

	Hiçbir Yan Etki Yok	Hafif	Ciddi	Satır Toplamı
ABD	90	45	15	150
Hindistan	75	65	15	150
Brezilya	80	50	20	150

Sütun Toplamı	245	155	50	450

Adım 1: Hipotezler ● H_0: Yan etkilerin dağılımı üç ülkede de aynıdır (nüfuslar homojendir) ● H_a: Yan etkilerin dağılımı ülkeler arasında farklılık gösterir

Adım 2: Beklenen Frekansları Hesaplayın

Bağımsızlık testi ile aynı formül:

"ABD + Yan Etki Yok" için: E = (150 × 245) / 450 = 81,67 "ABD + Hafif" için: E = (150 × 155) / 450 = 51,67 "ABD + Şiddetli" için: E = (150 × 50) / 450 = 16,67 (Ve bu şekilde 9 hücrenin tamamı için...)

Beklenen Tablo:

	Hiçbir Yan Etki Yok	Hafif	Ciddi
ABD	81.67	51.67	16.67
Hindistan	81.67	51.67	16.67
Brezilya	81.67	51.67	16.67

Adım 3: χ^2'yi hesaplayın

ABD, No: $(90 - 81{,}67)^2 / 81{,}67 = 0{,}85$ ABD, Hafif: $(45 - 51{,}67)^2 / 51{,}67 = 0{,}86$ ABD, Şiddetli: $(15 - 16{,}67)^2 / 16{,}67 = 0{,}17$ Hindistan, No: $(75 - 81{,}67)^2 / 81{,}67 = 0{,}54$ Hindistan, Hafif: $(60 - 51{,}67)^2 / 51{,}67 = 1{,}34$ Hindistan, Şiddetli: $(15 - 16{,}67)^2 / 16{,}67 = 0{,}17$ Brezilya, No: $(80 - 81{,}67)^2 / 81{,}67 = 0{,}03$ Brezilya, Hafif: $(50 - 51{,}67)^2 / 51{,}67 = 0{,}05$ Brezilya, Şiddetli: $(20 - 16{,}67)^2 / 16{,}67 = 0{,}67$

$\chi^2 = 0.85 + 0.86 + 0.17 + 0.54 + 1.34 + 0.17 + 0.03 + 0.05 + 0.67 = 4.68$

Adım 4: Serbestlik dereceleri df = (rows - 1) × (columns - 1) = (3 - 1) × (3 - 1) = 2 × 2 = 4

Adım 5: Karar

df = 4 ve $\alpha = 0{,}05$ için kritik değer = 9,49 4,68 < 9,49 olduğundan, H_0'yi reddedemeyiz

Sonuç: "Yan etki profili üç ülkede de benzer görünüyor. İlaç, hangi ülkede olursa olsun hastaları benzer şekilde etkiliyor gibi görünüyor."

In [4]:
```
import numpy as np
import pandas as pd
from scipy.stats import chi2_contingency

# Gözlenen veriler
observed = np.array([
    [90, 45, 15],      # ABD
    [75, 65, 15],      # Hindistan
    [80, 50, 20]       # Brezilya
])
```

```python
# Chi-square testi uygula
chi2_stat, p_value, dof, expected = chi2_contingency(observed, correction=False)

print("Chi-Square istatistiği:", round(chi2_stat, 3))
print("Serbestlik derecesi:", dof)
print("p-değeri:", round(p_value, 5))

# Beklenen değer tablosu
print("\nBeklenen Değerler:")
print(pd.DataFrame(np.round(expected, 2),
                   index=["ABD", "Hindistan", "Brezilya"],
                   columns=["Hiçbir Yan Etki Yok", "Hafif", "Ciddi"]))

# Karar
alpha = 0.05
if p_value < alpha:
    print("\nH0 reddedilir: Yan etkilerin dağılımı ülkeler arasında farklıdır.")
else:
    print("\nH0 reddedilemez: Yan etkilerin dağılımı ülkeler arasında aynıdır.")
```

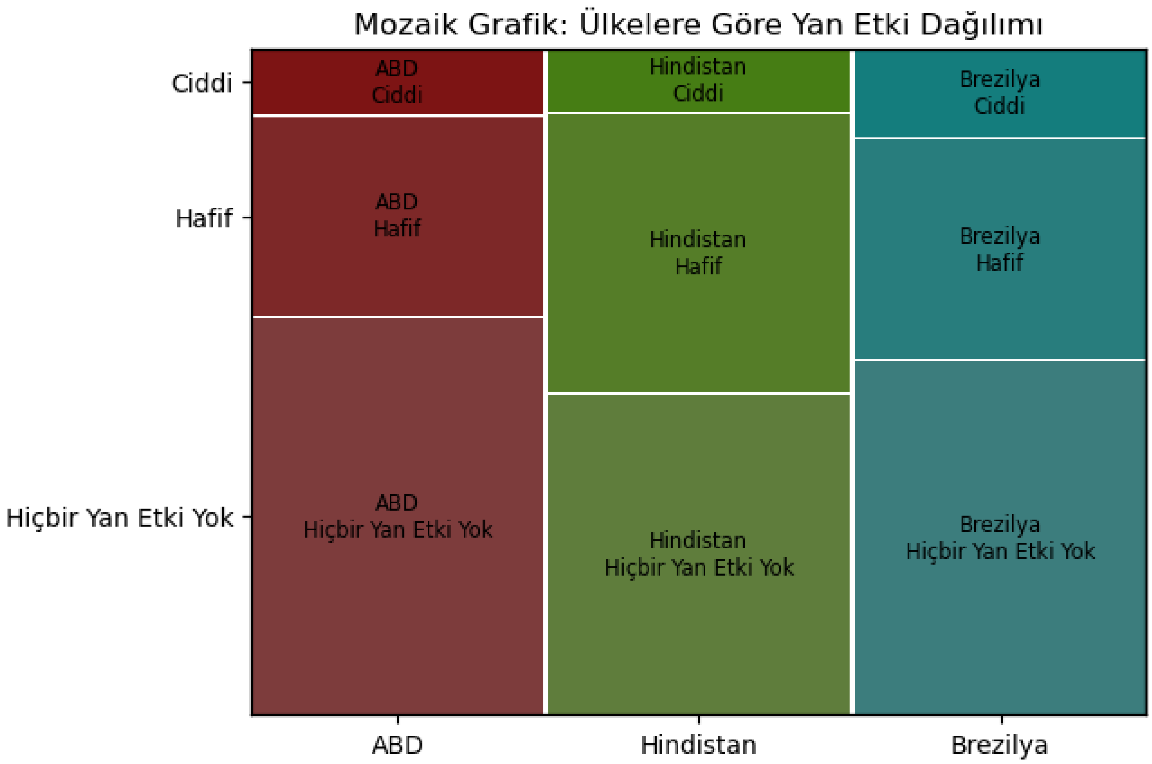
```
Chi-Square istatistiği: 6.348
Serbestlik derecesi: 4
p-değeri: 0.17462

Beklenen Değerler:
           Hiçbir Yan Etki Yok  Hafif  Ciddi
ABD                      80.77  52.75  16.48
Hindistan                83.46  54.51  17.03
Brezilya                 80.77  52.75  16.48

H0 reddedilemez: Yan etkilerin dağılımı ülkeler arasında aynıdır.
```

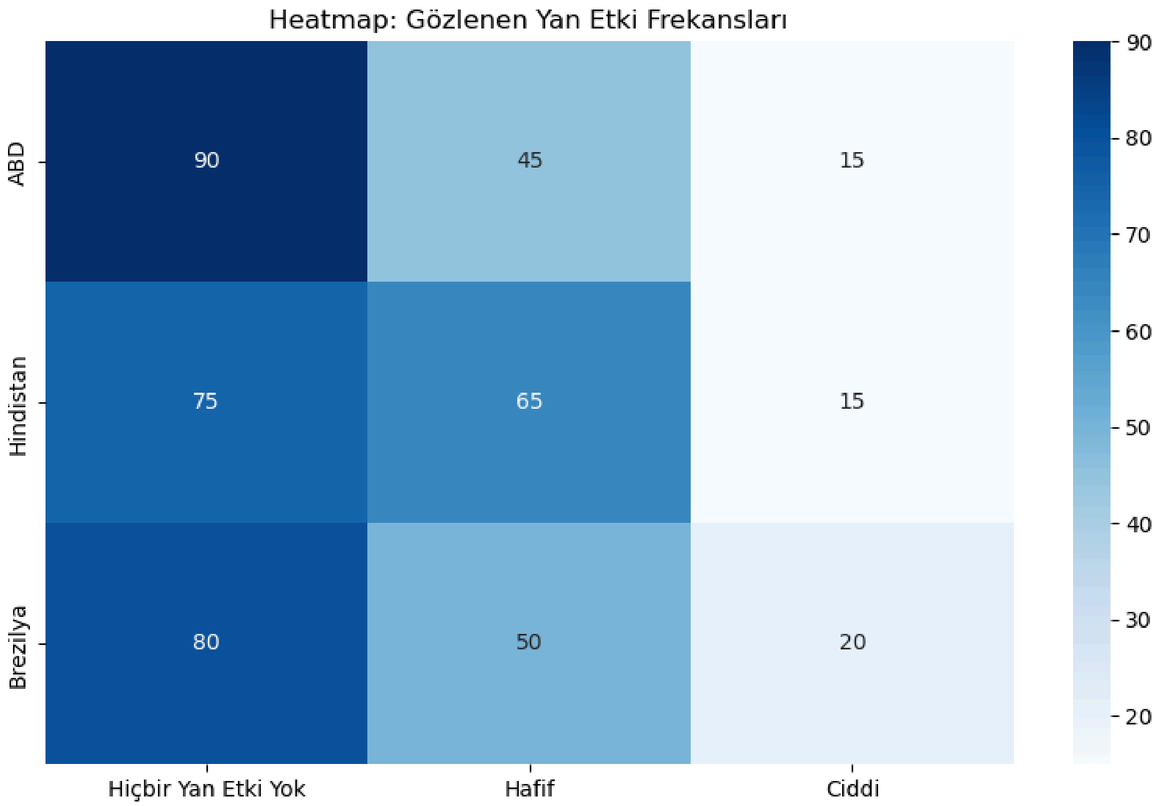
In [5]:

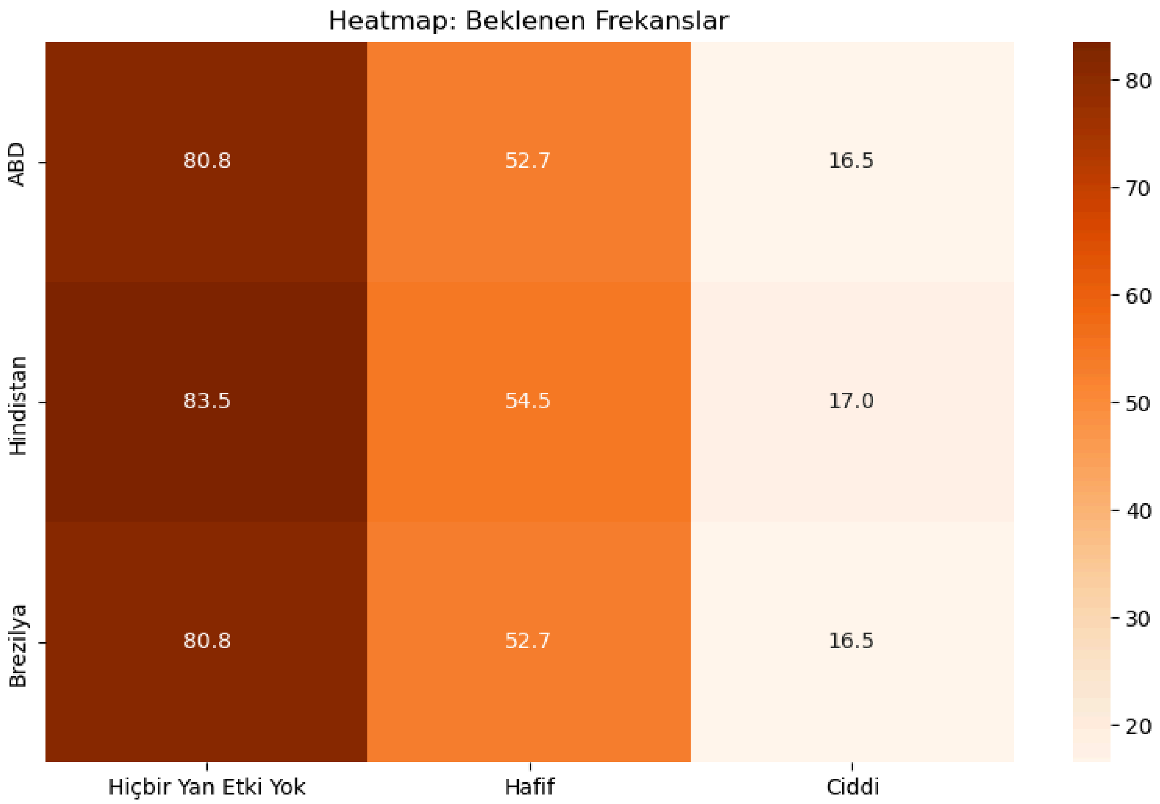
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.graphics.mosaicplot import mosaic
from scipy.stats import chi2_contingency

# Veri
observed = np.array([
    [90, 45, 15],
    [75, 65, 15],
    [80, 50, 20]
])

df = pd.DataFrame(observed,
                  index=["ABD", "Hindistan", "Brezilya"],
                  columns=["Hiçbir Yan Etki Yok", "Hafif", "Ciddi"])

# 🔹 Ki-Kare Testi
chi2_stat, p_value, dof, expected = chi2_contingency(observed, correction=False)

# ✅ Mozaik Grafik
plt.figure(figsize=(10, 6))
mosaic(df.stack())
plt.title("Mozaik Grafik: Ülkelere Göre Yan Etki Dağılımı")
plt.show()

# ✅ Heatmap (gözlenen değerler)
```

```
plt.figure(figsize=(10, 6))
sns.heatmap(df, annot=True, fmt="d", cmap="Blues")
plt.title("Heatmap: Gözlenen Yan Etki Frekansları")
plt.show()

# ✅ Heatmap (beklenen değerler)
expected_df = pd.DataFrame(expected,
                           index=df.index,
                           columns=df.columns)

plt.figure(figsize=(10, 6))
sns.heatmap(expected_df, annot=True, fmt=".1f", cmap="Oranges")
plt.title("Heatmap: Beklenen Frekanslar")
plt.show()
```

```
<Figure size 1000x600 with 0 Axes>
```







In [6]:

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

# Ham veri oluşturma
data = []

# ABD
data += [["ABD", "Hiçbir Yan Etki Yok"]] * 90
data += [["ABD", "Hafif"]] * 45
```

```
data += [["ABD", "Ciddi"]] * 15

# Hindistan
data += [["Hindistan", "Hiçbir Yan Etki Yok"]] * 75
data += [["Hindistan", "Hafif"]] * 65
data += [["Hindistan", "Ciddi"]] * 15

# Brezilya
data += [["Brezilya", "Hiçbir Yan Etki Yok"]] * 80
data += [["Brezilya", "Hafif"]] * 50
data += [["Brezilya", "Ciddi"]] * 20

# DataFrame
df = pd.DataFrame(data, columns=["Ülke", "Yan Etki"])

# Gözlenen çapraz tablo
contingency = pd.crosstab(df["Ülke"], df["Yan Etki"])
print("Gözlenen Tablo:")
print(contingency)

# Chi-square Test of Homogeneity
chi2_stat, p_value, dof, expected = chi2_contingency(contingency, correction=Fal

print("\nChi-Square İstatistiği:", round(chi2_stat, 3))
print("Serbestlik Derecesi:", dof)
print("p-değeri:", round(p_value, 5))

# Beklenen değerler tablosu
expected_df = pd.DataFrame(np.round(expected, 2),
                           index=contingency.index,
                           columns=contingency.columns)

print("\nBeklenen Değerler:")
print(expected_df)

# Karar
alpha = 0.05
if p_value < alpha:
    print("\nH0 reddedilir: Ülkeler arasında yan etki dağılımı farklıdır.")
else:
    print("\nH0 reddedilemez: Ülkeler arasında yan etki dağılımı benzerdir.")
```

```
Gözlenen Tablo:
Yan Etki   Ciddi  Hafif  Hiçbir Yan Etki Yok
Ülke
ABD           15     45                   90
Brezilya      20     50                   80
Hindistan     15     65                   75

Chi-Square İstatistiği: 6.348
Serbestlik Derecesi: 4
p-değeri: 0.17462

Beklenen Değerler:
Yan Etki   Ciddi  Hafif  Hiçbir Yan Etki Yok
Ülke
ABD        16.48  52.75                80.77
Brezilya   16.48  52.75                80.77
Hindistan  17.03  54.51                83.46

H0 reddedilemez: Ülkeler arasında yan etki dağılımı benzerdir.
```

import pandas as pd import numpy as np from scipy.stats import chi2

Ham veri oluşturma

data = []

ABD

data += [["ABD", "Hiçbir Yan Etki Yok"]] * 90 data += [["ABD", "Hafif"]] * 45 data += [["ABD", "Ciddi"]] * 15

Hindistan

data += [["Hindistan", "Hiçbir Yan Etki Yok"]] * 75 data += [["Hindistan", "Hafif"]] * 65 data += [["Hindistan", "Ciddi"]] * 15

Brezilya

data += [["Brezilya", "Hiçbir Yan Etki Yok"]] * 80 data += [["Brezilya", "Hafif"]] * 50 data += [["Brezilya", "Ciddi"]] * 20

df = pd.DataFrame(data, columns=["Ülke", "Yan Etki"])

Gözlenen tablo

observed = pd.crosstab(df["Ülke"], df["Yan Etki"]) print("📌 Gözlenen Tablo:") print(observed)

Satır ve sütun toplamları

row_totals = observed.sum(axis=1) col_totals = observed.sum(axis=0) grand_total = observed.values.sum()

Beklenen değerler

expected = np.outer(row_totals, col_totals) / grand_total expected_df = pd.DataFrame(np.round(expected, 2), index=observed.index, columns=observed.columns) print("\n📌 Beklenen Tablo:") print(expected_df)

✅ Chi-square Hesabı

chi_square = np.sum((observed.values - expected)**2 / expected)

Serbestlik derecesi

dfree = (observed.shape[0] - 1) * (observed.shape[1] - 1)

p-değeri

p_value = 1 - chi2.cdf(chi_square, dfree)

print(f"\n🔎 Chi-Square İstatistiği: {chi_square:.3f}") print(f"Serbestlik Derecesi: {dfree}") print(f"p-değeri: {p_value:.5f}")

Karar

alpha = 0.05 if p_value < alpha: print("\n❌ H0 reddedilir: Ülkeler arasında yan etki dağılımı farklıdır.") else: print("\n✅ H0 reddedilemez: Ülkeler arasında yan etki dağılımı benzerdir.")

Ki-Kare Testleri Hakkında Önemli Notlar

1. Örneklem Büyüklüğü Gereksinimleri: Ki-kare testleri makul derecede büyük beklenen frekanslar gerektirir. Genel kural:

- Beklenen TÜM frekanslar en az 5 olmalıdır

- Beklenen herhangi bir frekans 5'ten azsa, test güvenilir olmayabilir

Bu ihlal edildiğinde ne yapılmalı:

- Kategorileri birleştirin (mantıklıysa)
- Fisher's Exact Test kullanın (2×2 tablolar için)
- Daha fazla veri toplayın

2. Ki-Kare Size Sadece Bir İlişkinin Olup Olmadığını Söyler, Ne Kadar Güçlü Olduğunu Değil

Bağımsızlık testinde H_0'yi reddederseniz, değişkenlerin ilişkili olduğunu bilirsiniz, ancak ki-kare size ne kadar güçlü bir ilişki olduğunu göstermez. Bunun için Cramér'in V'si gibi ölçütlere bakmanız gerekir.

3. Yönün Önemi Yok T-testleri veya z-testlerinin aksine, ki-kare testleri özünde her zaman "çift taraflı"dır; sadece beklenenden HERHANGİ bir sapma olup olmadığını test edersiniz. Tek taraflı bir ki-kare testi yapamazsınız.

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```