

# Ahmet Burak Şişci

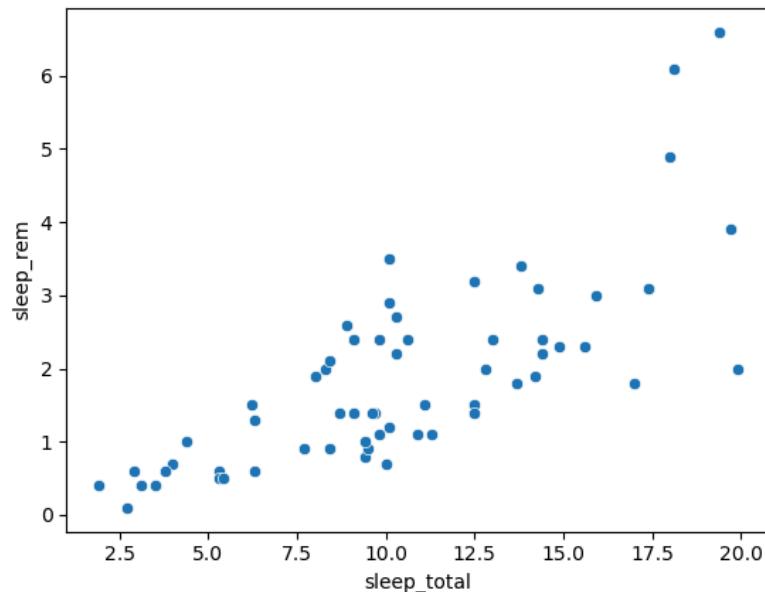
## sınav bilgileri

- boşluk doldurma ve yorum yok
  - bir şey veririm sonra onu yorumlamamızı isterim.
- düzungün yaz. - bence grafik verip yorum yapmamızı isteyebilir.
- yarı koda yarı puan yok. kod tam olacak.
  - defter, kitap açık.
  - cep telefonu toplanacak. başkasının notlarına bakmak yok.
  - herkes kendi notunu çıkaracak.

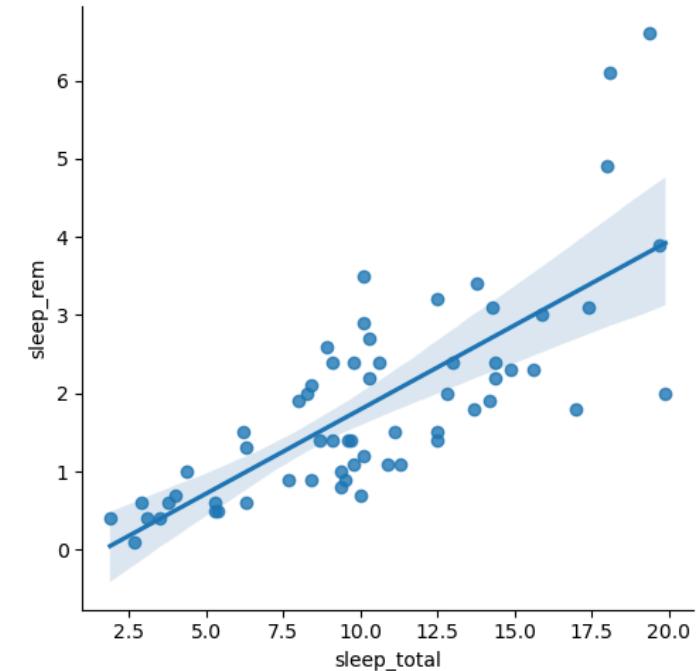
```
In [ ]: import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

df_sleep = pd.read_csv("msleep.csv")

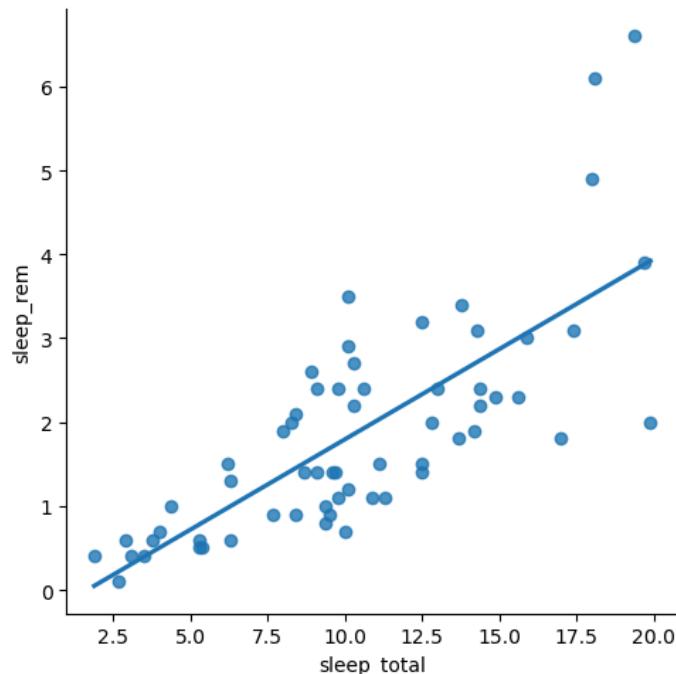
sns.scatterplot(x="sleep_total", y="sleep_rem", data=df_sleep)
plt.show()
```



```
In [ ]: import seaborn as sns
sns.lmplot(x="sleep_total", y="sleep_rem", data=df_sleep, ci=95)
plt.show()
```



```
In [ ]: import seaborn as sns
sns.lmplot(x="sleep_total", y="sleep_rem", data=df_sleep, ci=None) # ci = güven
plt.show()
```



`ci=None`, `lmplot` tarafından çizilen regresyon doğrusunun (uyum doğrusu) etrafındaki **Güven Aralığı (Confidence Interval)** gölgelerini kaldırma yarar.

- **ci Nedir?** "Confidence Interval" kelimesinin kısaltmasıdır.
- **Normalde Ne Yapar?** Eğer `ci` parametresini belirtmezseniz (`ci=95` varsayılan değerdir), `seaborn` regresyon doğrusunun etrafında soluk, şeffaf bir gölgeli alan çizer. Bu gölgelik, doğrunun gerçek konumunun %95 olasılıkla içinde bulunduğu aralığı temsil eder.
- **ci=None Ne İşe Yarar?** Bu gölgeli alanı **kapatır** ve sadece regresyon doğrusunun kendisini gösterir. Bu, grafiği daha sade ve temiz hale getirir.

**Özetle:** `ci=None`, "Regresyon doğrusunun etrafındaki belirsizlik gölgelerini gösterme, sadece en iyi uyum doğrusunu çiz" demektedir.

```
In [ ]: df_sleep["sleep_total"].corr(df_sleep["sleep_rem"])
```

```
Out[ ]: np.float64(0.7517549992287141)
```

Bu kod, `df_sleep` veri setindeki "**toplam uykı**" (`sleep_total`) ile "**REM uykusu**" (`sleep_rem`) sütunları arasındaki **Pearson Korelasyon Katsayısını** hesaplar.

Kısaltıcı, şu soruya cevaplar: "Bir hayvanın toplam uykı süresi artarken, REM uykusu süresi de artma veya azalma eğiliminde midir ve bu ilişki ne kadar güçlündür?"

## heatmap() fonksiyonu

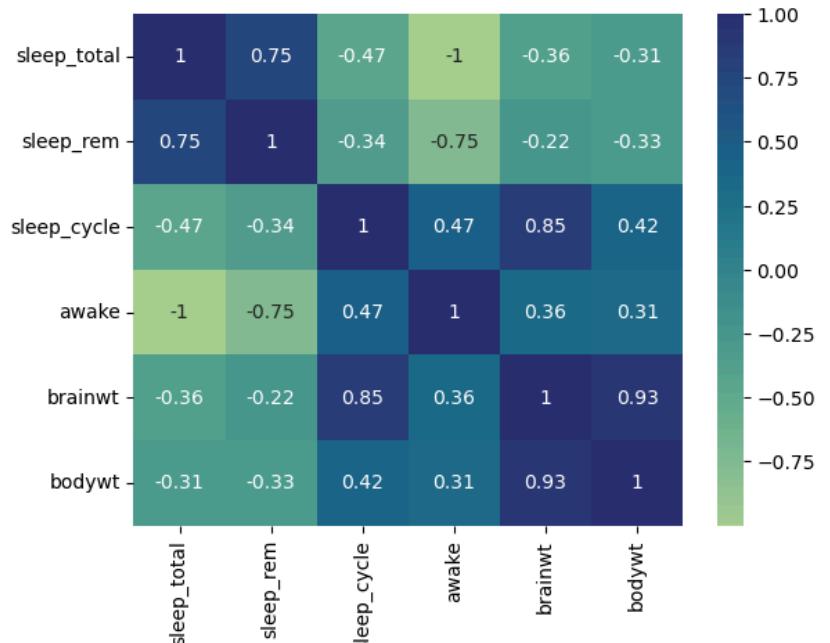
**Heatmap (Isı Haritası) Fonksiyonu**, bir tablodaki sayısal verileri, değerlerin büyüklüğüne göre **renklerle** görselleştiren bir grafik oluşturur.

- **Ne Yapar?** Sayısal bir matrisi (tabloyu) alır ve her bir hücreyi, içindeki sayının değerine göre renklendirir.
- **Amacı Nedir?** Tablodaki **desenleri, yoğunlukları ve aynırı değerleri** bir bakışta hızla fark etmemizi sağlar.
- **Renkler Ne Anlama Gelir?** Genellikle,
  - **Koyu renkler** (veya sıcak renkler: kırmızı, turuncu) → **Yüksek değerleri**
  - **Açık renkler** (veya soğuk renkler: mavi, yeşil) → **Düşük değerleri**
 gösterir.

**Özetle:** Sayılarla dolu karmaşık bir tabloyu, anlaşılması kolay bir renkli resme dönüştürür. Özellikle korelasyon matrislerini görselleştirmek için çok sık kullanılır.

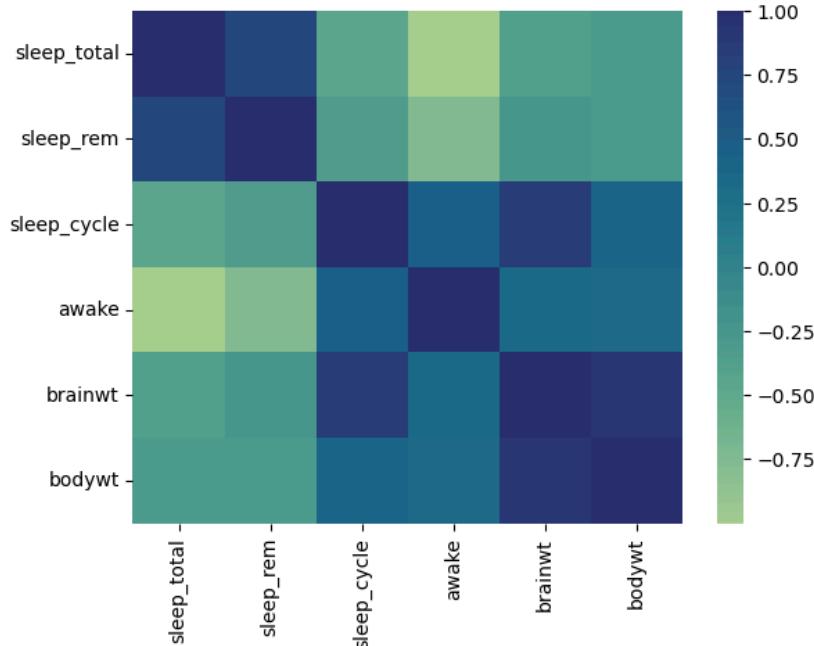
```
In [ ]: sns.heatmap(df_sleep.corr(numeric_only=True), annot=True, cmap='crest') # cmap kullanmak zorunda değilsin
```

```
Out[ ]: <Axes: >
```



```
In [ ]: sns.heatmap(df_sleep.corr(numeric_only=True), annot=False, # annot üzerinde yazı cmap='crest') # cmap kullanmak zorunda değilsin
```

```
Out[ ]: <Axes: >
```



## Kodun Kısa Açıklaması

Bu kod, `df_sleep` veri setindeki **tüm sayısal sütunların** birbiriley olan **korelasyon matrisini** hesaplar ve bu matrisi bir **ısı haritası (heatmap)** olarak görselleştirir.

- `df_sleep.corr(numeric_only=True)` : `df_sleep` içindeki sadece sayısal sütunlar arasındaki korelasyonları (-1 ile +1 arası değerler) hesaplayıp bir tablo (matris) oluşturur.
- `sns.heatmap(...)` : Bu korelasyon tablosunu alır ve renkli bir grafiğe dönüştürür.
- `annot=True` : Her bir renkli karenin içine, o hücreye ait olan korelasyon değerini (sayılı) yazar. Bu, grafiği daha okunaklı hale getirir.

**Özetle:** "Hangi iki değişken arasında ne kadar güçlü bir ilişki var?" sorusunu, renkli ve üzerinde sayılar yazan bir tablo ile görsel olarak cevaplar.

## cmap Nedir? Kısaca

`cmap`, "colormap" (renk haritası) kelimesinin kısaltmasıdır.

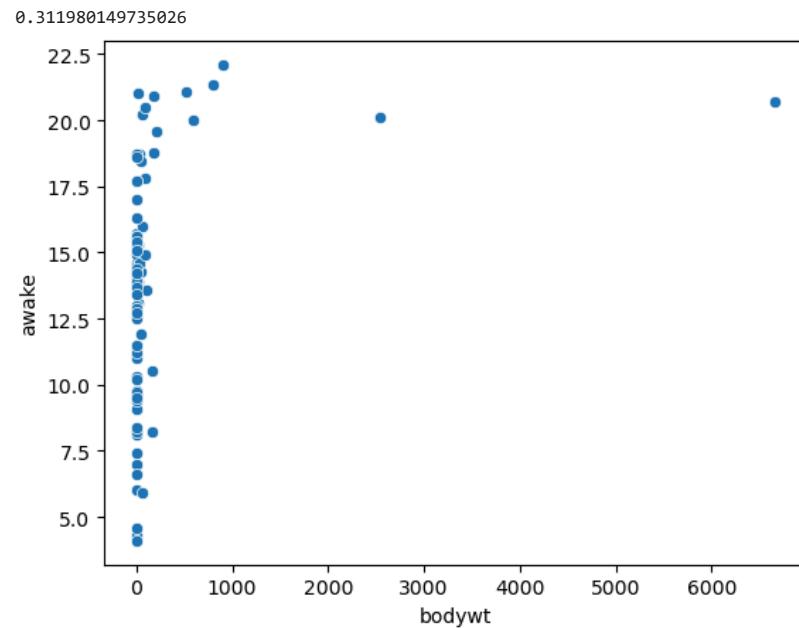
Isı haritasında kullanılacak **renk paletini** veya **renk şemasını** belirler.

- **Ne İşe Yarar?** Düşük değerlerden yüksek değerlere geçişte hangi renklerin kullanılacağını tanımlar.

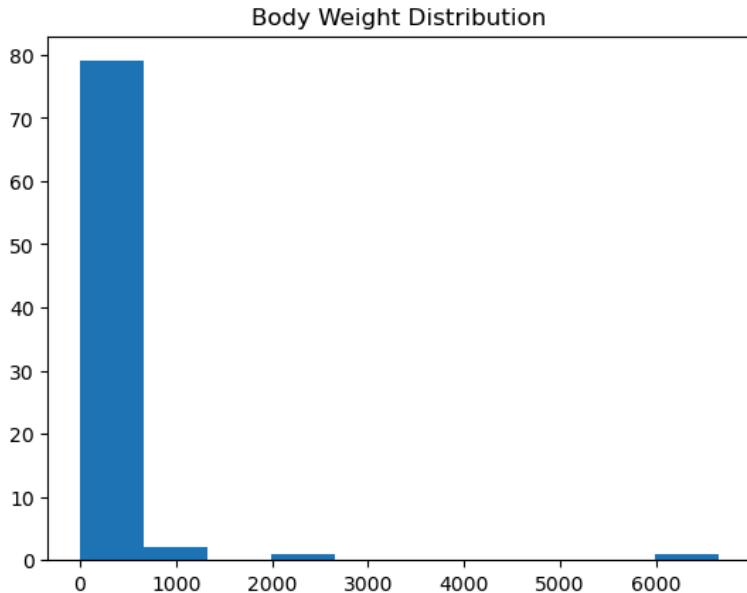
- `cmap='crest'` : Bu örnekte, `seaborn`'un hazır paletlerinden biri olan "crest" kullanılır. Bu palet genellikle düşük değerler için açık yeşil/mavi tonlarından başlayıp yüksek değerler için koyu deniz mavisi/mor tonlarına doğru giden bir renk geçisi sunar.

bazen (lineer olmayan - düz olmayan) grafiklerde korelasyon sayısı düşük çıkabilir. bunu için grafiğe bakmak gereklidir.

```
In [ ]: import seaborn as sns
sns.scatterplot(x="bodywt", y="awake", data=df_sleep)
print(df_sleep["bodywt"].corr(df_sleep["awake"]))
```

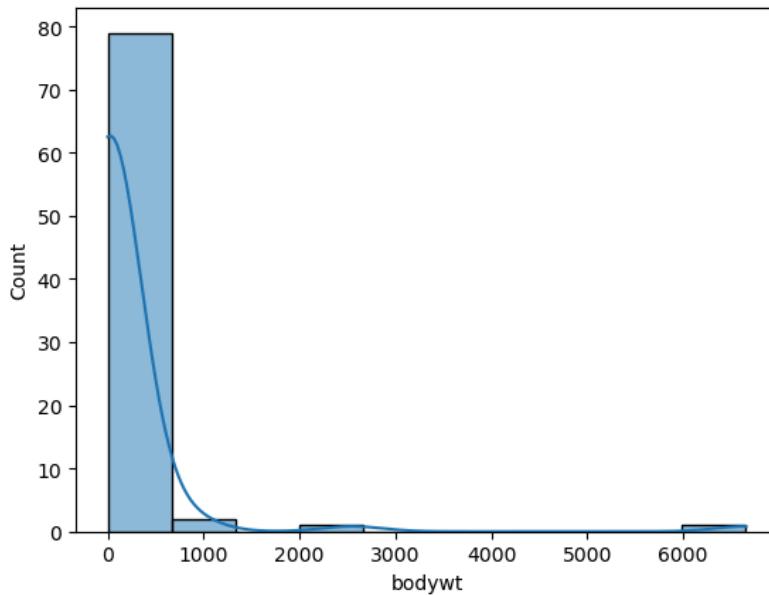


```
In [ ]: plt.hist(df_sleep["bodywt"])
plt.title("Body Weight Distribution")
plt.show()
```



```
In [ ]: sns.histplot(data=df_sleep, x='bodywt', kde=True, bins=10)
```

```
Out[ ]: <Axes: xlabel='bodywt', ylabel='Count'>
```



1. `plt.hist(df_sleep["bodywt"])` (Matplotlib Grafiği)

- **Ne Yapar?** `df_sleep` veri setindeki `bodywt` (vücut ağırlığı) sütununun **temel bir histogramını** çizer.
- **Yorumu:** Bu grafik, vücut ağırlığı verilerinin dağılımını gösterir. Büyük ihtimalle, verilerin büyük bir çoğunluğu grafiğin sol tarafında (düşük ağırlık değerlerinde) toplanmıştır ve sağa doğru uzanan uzun bir kuyruk vardır. Bu, dağılımin **sağa çarpık (positively skewed)** olduğunu gösterir. Yani, çoğu hayvan düşük kiloluyken, çok az sayıda aşırı ağır hayvan bulunmaktadır.

## 2. `sns.histplot(..., kde=True, bins=10)` (Seaborn Grafiği)

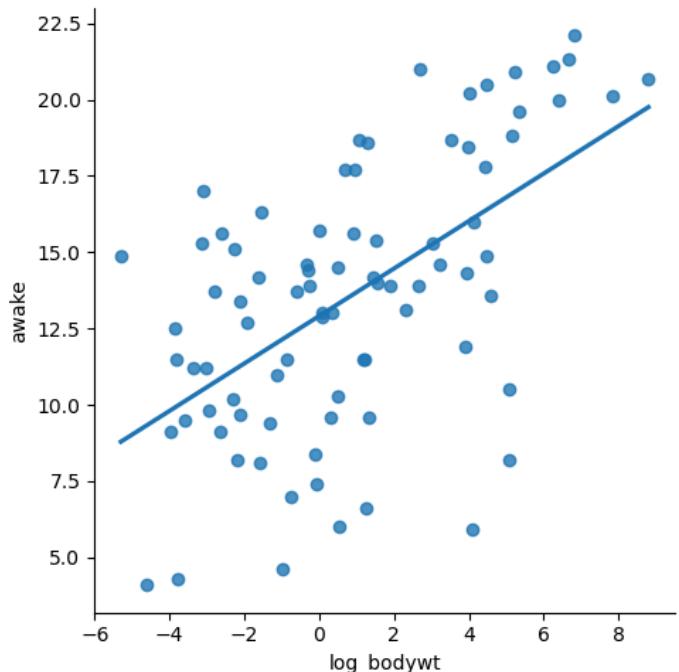
- **Ne Yapar?** Matplotlib ile aynı histogramı çizer, ancak birkaç ek özellik sunar:
  - **Daha Estetik:** seaborn 'un varsayılan stilleri sayesinde genellikle daha modern ve okunaklı bir görünüm sunar.
  - **kde=True :** Histogramın üzerinde, verinin olasılık yoğunluğunu tahmin eden pürüzsüz bir eğri (**Kernel Density Estimate - KDE**) çizer. Bu eğri, dağılımin genel şeklini daha net görmemizi sağlar.
  - **bins=10 :** Veri aralığının kaç adet cubuğa (bölmeye) ayrılacağını manuel olarak belirler.
- **Yorumu:** Bu grafik de vücut ağırlığının **sağa çarpık** olduğunu teyit eder. Ek olarak çizilen KDE eğrisi, bu çarpıklığı ve tepe noktasının nerede yoğunlaştığını daha pürüzsüz bir şekilde vurgular. `bins=10` parametresi, grafiğin detay seviyesini ayarlar; daha az veya daha çok çubukla dağılımin farklı bir yorumu elde edilebilir.

**Özetle:** Her iki grafik de aynı temel bilgiyi (düzenli bir sağa çarpık dağılımını) gösterir. Ancak `seaborn` grafiği, `kde` eğrisi sayesinde dağılımin şeklini daha net bir şekilde özetler ve genellikle görsel olarak daha çekicidir.

```
In [ ]: import numpy as np
df_sleep["log_bodywt"] = np.log(df_sleep["bodywt"])
print(df_sleep["log_bodywt"].corr(df_sleep["awake"]))

sns.lmplot(x="log_bodywt", y="awake", data=df_sleep, ci=None)
plt.show()
print()
```

```
0.5687943427609857
```



- elimde çok çarpık veriler varsa log dönüşümü yapılabilir.

### Log Dönüşümü Nedir? (Kısa Tanım)

**Log Dönüşümü**, bir veri setindeki her bir değerin **logaritmasını** alarak verileri **yeniden ölçeklendirme** işlemidir.

Kısacası, elinizdeki  $[10, 100, 1000]$  gibi bir veri setini, logaritmasını alarak  $[1, 2, 3]$  (10 tabanında) gibi yeni bir sete dönüştürmektedir.

### Mantığı ve Amacı Nedir? Neden Yapılır?

Log dönüşümünün temel amacı, **sağa çarlık (positively skewed)** verileri daha **simetrik** ve **Normal Dağılım'a benzer** bir hale getirmektir.

Bunu neden yaparız?

#### 1. Büyük Değerlerin Etkisini Azaltmak:

- Sorun:** Bir veri setinde  $[1, 2, 5, 10, 1000]$  gibi değerler varsa,  $1000$  gibi aşırı büyük bir değer (aykırı değer), ortalama ve standart sapma gibi istatistikleri ve birçok grafiği domine eder. Diğer küçük değerlerin arasındaki ilişkiyi görmek zorlaşır.

- Çözüm:** Logaritma, büyük sayıları "küçültürken" küçük sayıları daha az etkiler.  $\log_{10}$  dönüşümü sonrası veri seti  $[0, 0.3, 0.7, 1, 3]$  gibi olur. Artık  $3$  değeri, diğerlerini ezmeden, daha yönetilebilir bir ölçektektir.

#### 2. İlişkileri Doğrusallaştırmak (Lineer Hale Getirmek):

- Bazen iki değişken arasındaki ilişki üstel (eğrisel) olabilir. Bu ilişkinin logaritmmasını almak, onu genellikle düz bir çizgiye (doğrusal bir ilişkiye) dönüştürür. Bu, doğrusal regresyon gibi modellerin daha iyi çalışmasını sağlar.

#### 3. İstatistiksel Varsayımları Sağlamak:

- Birçok istatistiksel test (t-testi, ANOVA, doğrusal regresyon vb.), verilerin Normal Dağılım'a uymasını varsayar. Eğer verileriniz sağa çarpıksa, bu varsayımlı ihlal edilir. Log dönüşümü, veriyi normale yaklaştırarak bu testlerin daha güvenilir sonuçlar vermesini sağlar.

### Hangi Durumlarda Kullanılır?

- Sağça Çarlık Verilerde:** Gelir dağılımı, ev fiyatları, bir web sitesindeki sayfa ziyaret sayıları, bakteri popülasyonları gibi veriler genellikle sağa çarptır.
- Çarpımsal İlişkilerde:** Etkilerin toplamsal değil, çarpımsal olduğu durumlarda (örneğin, biyolojik büyümeye süreçleri).
- Veri Aralığı Çok Geniş Olduğuunda:** Milyonlarla ifade edilen değerlerle birlikte onlarla ifade edilen değerler varsa.

### Özet

Log dönüşümü, verilerinizi **"farklı bir gözlükle"** görmenizi sağlayan bir araçtır. Aşırı büyük sayıların baskın etkisini azaltarak, verinin altında yatan gerçek deseni ve ilişkileri daha net bir şekilde ortaya çıkarır. Bu, hem görselleştirmeyi iyileştirir hem de birçok istatistiksel modelin daha doğru çalışmasını sağlar.

**İkik değişken yüksek kolerasyona sahip olsada bu kesinlikle nedensellik olduğu anlamına gelmez.**

**İkisi arasında ilişki olabilir ama nedensellik olduğu anlamına gelmez**

**Sahte Korelasyon (Spurious Correlation)**, iki değişken arasında matematiksel olarak bir ilişki (korelasyon) varmış gibi görünmesi, ancak aslında bu ilişkinin tamamen **tesadüf** eseri olması veya her ikisini de etkileyen **gizli bir üçüncü değişkenden** kaynaklanması durumudur.

**Kıcacısı:** İki şey birlikte hareket ediyor gibi görünür, ama aralarında mantıklı bir **neden-sonuç ilişkisi yoktur**.

**Klasik Örnek:** Dondurma satışları ile boğulma vakaları arasında güçlü bir pozitif korelasyon vardır. Ancak dondurma yemek boğulmaya neden olmaz. Her ikisini de artıran gizli üçüncü değişken **sıcak havadır**.

## kahve ile akciğer kanseri arasındaki ilişki gibi

çünkü çok sigara içen çok kahve içen ve çok sigara içen kanser olur. sigara-kanser ilişkisinde nedensellik var ama kahve-kanserde nedensellik yok. ilişki var.

not: kahveyi severim. <3

**Anlamlılık Düzeyi (Significance Level -  $\alpha$ )**, bir hipotez testinde, aslında doğru olan bir **sıfır hipotezini ( $H_0$ ) yanlışlıkla reddetme riskidir.**

Kısacası, "tesadüfen anlamlı bir sonuç bulup **yanılma payım ne kadar olsun?**" sorusunun cevabıdır.

- **Ne İşe Yarar?** Bir sonucun "istatistiksel olarak anlamlı" olup olmadığına karar vermek için bir **eşik değeri** belirler.
- **Yaygın Değerler:** Genellikle **%5 ( $\alpha = 0.05$ )** olarak seçilir. Bu, bulduğumuz sonucun tesadüf olma ihtimalini **5%**'in altına düşürmeyi hedeflediğimiz anlamına gelir. Bazen **%1 ( $\alpha = 0.01$ )** veya **%10 ( $\alpha = 0.10$ )** da kullanılır.
- **Karar Kuralı:** Eğer testten elde edilen **p-değeri**, belirlediğimiz anlamlılık düzeyinden **küçükse** ( $p < \alpha$ ), sonuç "istatistiksel olarak anlamlı" kabul edilir ve sıfır hipotezi reddedilir.

## tip 1 hata ve tip 2 hata nedir (mahkeme anolojisi)

Bir sanığın durumu:

- **Sıfır Hipotezi ( $H_0$ ):** Sanık masumdur.

### Tip 1 Hata ( $\alpha$ ) - "Masumu Asmak"

- **Nedir?** Aslında **doğru** olan bir sıfır hipotezini **yanlışlıkla reddetmek**.
- **Analoji:** Gerçekte **masum** olan bir sanığı, **suçlu** bulmaktır.
- **Diğer Adı:** Yanlış Pozitif (False Positive).
- **Kontrolü:** Bu hatayı yapma olasılığı, bizim belirlediğimiz **anlamlılık düzeyine ( $\alpha$ )** eşittir.  $\alpha$ 'yı düşürerek bu riski azaltabiliriz.

### Tip 2 Hata ( $\beta$ ) - "Suçluyu Salıvermek"

- **Nedir?** Aslında **yanlış** olan bir sıfır hipotezini **yanlışlıkla kabul etmek**.
- **Analoji:** Gerçekte **suçlu** olan bir sanığı, **masum** bulup serbest bırakmaktadır.
- **Diğer Adı:** Yanlış Negatif (False Negative).
- **Neden Olur?** Genellikle örneklem boyutunun küçük olması veya etkinin zayıf olması gibi nedenlerle, var olan bir farkı veya ilişkiye tespit edemediğimizde ortaya çıkar.

### Özetle:

- **Tip 1:** Olmayan bir şeyi "var" sanmak.
- **Tip 2:** Var olan bir şeyi "yok" sanmak.

## hoca az önce bu slayıtı bizimle paylaşacağını ilettili.

bizim konumuz, verilerden anlamlı sonuçlar çıkarmak ve bu sonuçlara dayanarak kararlar vermektedir. Hipotez testleri bu sürecin merkezindedir.

**Sıfır Hipotezi ( $H_0$ ):** Genellikle "hiçbir etki yoktur" veya "iki grup arasında fark yoktur" şeklinde kurulan varsayımdır.

- **Örnek  $H_0$ :** Yeni bir web sitesi tasarımının, eski tasarıma göre satışları artırmada **hiçbir etkisi yoktur**.

**Alternatif Hipotez ( $H_1$ ):** Araştırmacının kanıtlamaya çalıştığı şeydir.

- **Örnek  $H_1$ :** Yeni web sitesi tasarımları, satışları **artırır**.

İşte Tip 1 ve Tip 2 hataların bizim konumuzla doğrudan alakası:

### Tip 1 Hata'nın Konumuzla Alakası ("Masumu Asmak")

- **Senaryo:** Yaptığımız analiz sonucunda, yeni web sitesi tasarımlının satışları "istatistiksel olarak anlamlı" bir şekilde artıldığına karar verdik ( $p < 0.05$ ). Sıfır hipotezini reddettik.
- **Tip 1 Hata Durumu:** Eğer bu sonuç **tamamen şans eseri** ortaya çıktıysa ve gerçekte yeni tasarımın hiçbir etkisi yoksa, bir Tip 1 hata yapmış oluruz.
- **Sonucu (Bizim İçin):**
  - Şirket, aslında işe yaramayan yeni bir tasarıma **boşuna para ve zaman harcar**.
  - Yanlış bir stratejiye yatırım yapılır.
  - Analizimizin **güvenilirliği sarsılır**.

Kısacası, Tip 1 hata, bizi yanlış bir aksiyon almaya yönlendiren "yanlış alarmdır".

### Tip 2 Hata'nın Konumuzla Alakası ("Suçluyu Salıvermek")

- **Senaryo:** Yaptığımız analiz sonucunda, yeni tasarımın etkisinin "istatistiksel olarak anlamlı olmadığını" karar verdik ( $p > 0.05$ ). Sıfır hipotezini reddedemedik.
- **Tip 2 Hata Durumu:** Eğer yeni tasarım **gerçekten de işe yarıyorsa**, ama bizim örneklemimiz bu etkiyi tespit edecek kadar büyük veya güçlü değilse, bir Tip 2 hata

yapmış oluruz.

- **Sonucu (Bizim İçin):**

- Şirket, satışları artırabilecek **öneMLİ bir fırsatı kaçıRır**.
- İşe yarar bir inovasyon, "etkisiz" damgası iyiyerek rafa kaldırılır.
- Potansiyel bir rekabet avantajı kaybedilir.

Kısacası, Tip 2 hata, bizi kârlı veya faydalı bir aksiyon almaktan alıkoyan "kaçırılmış fırsat".

## Özet ve Bağlantı

Veri analistleri olarak amacımız sadece bir **p-değeri** hesaplamak değil, bu **p-değerini** yorumlayarak doğru iş kararları alınmasını sağlamaktır.

- **Anlamlılık Düzeyini ( $\alpha$ )** belirlerken, aslında ne kadar **Tip 1 hata riski** almayı göze alındığımıza karar veririz.
- Bir testin **istatistiksel gücünü (Power = 1 -  $\beta$ )** hesaplarken, ne kadar **Tip 2 hata riski** taşıdığımızı anlamaya çalışırız.

Bu iki hata türü arasında bir denge vardır. Tip 1 hata riskini çok azaltırsanız (örneğin  $\alpha$ 'yı 0.001 yaparsanız), Tip 2 hata yapma riskiniz artar (çünkü bir etkiye "anlamlı" kabul etmek için çok daha güçlü kanıtlar istersiniz).

Bu nedenle, yaptığımız her hipotez testinin sonucunu yorumlarken, "Bu sonuç ne kadar güvenilir? Yanlış alarm olma ihtimali nedir? Gerçek bir etkiye kaçırıyor olma ihtimalimiz nedir?" gibi soruları sormak zorundayız. Tip 1 ve Tip 2 hatalar, bu soruların istatistiksel dilindeki karşılığıdır.

## Hipotez Testine Giriş ve Hipotez Testinin Temelleri

bir iddayı, çıkarımı ispatlama zorunluluğu diyebilir. o şey şans eseri mi gelişti yoksa bizim tespit ettiğimiz nedene bağlı mı gelişti gibi bir durum

kısaca işini şansa bırakmama sanatı diyelim

### Hipotez Testine Giriş: Temel Fikir

**Hipotez Testi**, bir iddiayı veya varsayıyı, elimizdeki verileri kullanarak **sistematiK bir şekilde test etme** sürecidir.

**Amacı:** Gözlemlediğimiz bir etkinin veya farkın, **gerçek bir durumdan mı kaynaklandığını**, yoksa sadece **şans eseri (tesadüfen)** mi ortaya çıktığını anlamaktır.

## Hipotez Testinin Temelleri: 4 Adımlık Süreç

Bir dedektifin bir davayı çözmesi gibi, hipotez testi de belirli adımları takip eder.

### Adım 1: İki Farklı Senaryo Belirle (Hipotezleri Kur)

Bir dedektif, "Sanık ya suçludur ya da masumdur" der. Biz de iki zıt senaryo kurarız:

#### 1. Sıfır Hipotezi ( $H_0$ - Null Hypothesis):

- **Nedir?** Genellikle "hiçbir etki yok", "fark yok" veya "ilişki yok" şeklinde kurulan, **statükoYU savunan** varsayımdır.
- **Mantiği:** Bu, bizim **cürütmeye çalışacağımız** şüphelinin masumiyet karunesidir. Varsayılan olarak doğru kabul edilir.
- **Örnek:** "Yeni geliştirilen ilaçın, mevcut ilaca göre iyileşme süresine **hiçbir etkisi yoktur**."

#### 2. Alternatif Hipotez ( $H_1$ veya $H_a$ - Alternative Hypothesis):

- **Nedir?** Araştırmacının **kanıtlamaya çalıştığı** iddiadır. "Bir etki var", "fark var" veya "ilişki var" der.
- **Mantiği:** Bu, dedektifin "Sanık suçludur" iddiasıdır.
- **Örnek:** "Yeni ilaç, iyileşme süresini **kısaltır**."

### Adım 2: Kanıt Topla (Veri ve Test İstatistiği)

Dedektif, olay yerinden kanıt toplar. Biz de popülasyondan bir **örneklem alır** ve bu veriyi özetleyen tek bir sayı, yani bir **test istatistiği** hesaplarız.

- **Test İstatistiği:** Elimizdeki verinin, Sıfır Hipotezi'nin doğru olduğu varsayımi altında ne kadar "aşırı" veya "beklenmedik" olduğunu ölçen bir sayıdır. (Örn: t-skoru, z-skoru, ki-kare değerleri).
- **Mantiği:** Bu, olay yerinde bulunan ve sanığı işaret eden bir kanıt (örneğin, bir parmak izi) gibidir.

### Adım 3: Kanıtı Değerlendir (p-değeri)

Dedektif, "Eğer sanık masum olsaydı, bu parmak izinin burada bulunma ihtimali ne olurdu?" diye sorar. Biz de benzer bir soru sorarız:

- **p-değeri (p-value):** "Eğer Sıfır Hipotezi ( $H_0$ ) doğru olsaydı, en az bu kadar aşırı veya beklenmedik bir test istatistiği (kanıt) elde etme olasılığımız ne olurdu?"

### p-değerini Yorumlama:

- **Küçük p-değeri (örn:  $p < 0.05$ ):** "Eğer sanık masum olsaydı, bu kanıtın burada bulunması çok ama çok düşük bir ihtimaldi. Bu durum çok şüpheli!"
- **Büyük p-değeri (örn:  $p > 0.05$ ):** "Eğer sanık masum olsaydı, bu kanıtın burada bulunması gayet olası bir durum. Bu kanıt tek başına bir şey ifade etmiyor."

### Adım 4: Karar Ver (Sonuca Ulaş)

Dedektif, kanıtın ne kadar güçlü olduğuna bakarak bir karara varır. Biz de p-değerini, daha önceden belirlediğimiz bir **eşik değeri** ile karşılaştırırız.

- Anlamlılık Düzeyi ( $\alpha$  - Significance Level):** Bu, bizim "şüphe eşliğimizdir". Genellikle  $\alpha = 0.05$  (%5) olarak belirlenir. Bu, %5'ten daha düşük bir olasılıkla gerçekleşen olayları "şüpheli" ve "tesadüf değil" olarak kabul edeceğimiz anlamına gelir.

#### Karar Kuralı:

- Eğer  $p < \alpha$  ise (örn:  $0.02 < 0.05$ ):

- Sonuç **istatistiksel olarak anlamlıdır**.
- Kanıtımız,  $H_0$ 'in doğru olması durumunda beklenmeyecek kadar güçlüdür.
- Sıfır Hipotezini ( $H_0$ ) reddederiz.**
- Dedektifin Kararı:* "Sanığın masum olma ihtimali çok düşük, bu yüzden onu suçlu buluyorum." (Alternatif hipotezi destekleriz).

- Eğer  $p \geq \alpha$  ise (örn:  $0.34 > 0.05$ ):

- Sonuç istatistiksel olarak anlamlı **değildir**.
- Kanıtımız yeterince güçlü değil. Gözlemlediğimiz etki, şans eseri ortaya çıkmış olabilir.
- Sıfır Hipotezini ( $H_0$ ) reddedemeyiz.** (Dikkat: "Kabul ederiz" demiyoruz, sadece "reddedemeyiz" diyoruz).
- Dedektifin Kararı:* "Sanığın suçu olduğuna dair yeterli kanıt yok, bu yüzden onu masum kabul etmeye devam ediyoruz."

## Özet

Hipotez testi, bir iddiayı şüpheyele karşılayıp, verİYE dayalı kanıtlar işliğinde bu şüphenin ne kadar haklı olduğunu objektif bir şekilde ölçme sürecidir. Bu süreç, belirsizliklerle dolu dünyada daha bilimsel ve güvenilir kararlar almamızı sağlar.

## İstatistiksel çıkarım: küçük bir örneklem ile tüm bir grup hakkında sonuca varmaktadır

- örneğin: tenceereden 1 kaşık çorba alıp tüm tencerelerdeki çorbanın tuzuna bakmak.

**Sıfır Hipotezi ( $H_0$ )**, bir hipotez testinde, "iki grup arasında **fark yoktur**", "bir tedavinin **etkisi yoktur**" veya "iki değişken arasında **ilişki yoktur**" şeklinde kurulan, **varsayılan olarak doğru kabul edilen** ve testin **çürütmeye çalıştığı** başlangıç varsayımasıdır.

Kısacası, "değişik bir durum yok, her şey normal" diyen iddiadır.

**Alternatif Hipotez ( $H_1$  veya  $H_a$ )**, bir hipotez testinde, "bir **fark vardır**", "bir **etki vardır**" veya "bir **ilişki vardır**" şeklinde kurulan, araştırmacının **kanıtlamaya çalıştığı** ve Sıfır Hipotezi'nin ( $H_0$ ) reddedilmesi durumunda kabul edilecek olan iddiadır.

Kısacası, "değişik bir durum var, bir şeyler oluyor" diyen iddiadır.

## Mahkeme analogisi, hipotez testinin mantığını anlamak için mükemmel bir örnektir.

### Senaryo: Bir Cinayet Davası

Bir mahkeme salonundasınız ve bir sanık cinayete yargılanıyor. Hukuk sisteminin temel ilkesi "suçu ispatlanana kadar herkes masumdur".

Bu durumu hipotez testi diline çevirelim:

#### Adım 1: Hipotezleri Kurma

- Sıfır Hipotezi ( $H_0$ ): Sanık Masumdur.**

- Bu, varsayılan durumdur. Mahkeme, davaya bu varsayımla başlar. Savcının görevi, bu hipotezi makul bir şüphenin ötesinde çürütmektir.

- Alternatif Hipotez ( $H_1$ ): Sanık Suçludur.**

- Bu, savcının kanıtlamaya çalıştığı iddiadır.

#### Adım 2: Kanıt Toplama

- Savcı (Araştırmacı):** Olay yerinden kanıtlar toplar.

- Cinayet silahında sanığın parmak izleri.
- Güvenlik kamerası görüntüleri.
- Tanık ifadeleri.

- İstatistiksel Karşılığı:** Bu kanıtlar, bizim veri setimizden hesapladığımız **test istatistiğidir**.

#### Adım 3: Kanıtı Değerlendirme (p-değeri)

- Jüri (Karar Mekanizması):** Kendi kendine şu soruya sorar:

- "**Eğer sanık gerçekten masum olsaydı ( $H_0$  doğru olsaydı)**, tüm bu kanıtların (parmak izi, kamera kaydı vb.) bir araya gelme olasılığı ne olurdu?"

- Bu sorunun cevabı, p-değeri**dir.

- p-değeri çok küçükse:** "Masum birinin bu kadar çok kanıt bırakması neredeyse imkansız. Bu durum çok şüpheli."

- p-değeri büyükse:** "Bu kanıtlar zayıf. Masum biri de tesadüfen orada bulunmuş olabilir. Yeterince ikna edici değil."

#### Adım 4: Karar Verme

Jüri, kanıtların "makul bir şüphenin ötesinde" olup olmadığına karar verir. Bu, bizim **anlamlılık düzeyimizdir ( $\alpha$ )**.

- **Karar 1: Suçlu Bulma ( $H_0$ 'ı Reddetme)**

- **Durum:** Kanıtlar o kadar güclüdür ki, masum birinin bu kanıtları bırakma olasılığı aşırı düşüktür ( $p < \alpha$ ).
- **Jüri Kararı:** "Sıfır hipotezini (sanık masumdur) reddediyoruz. Sanığı suçlu buluyoruz."
- **Hata Riski:** Eğer sanık gerçekten masumsa ama kanıtlar şans eseri onu işaret ediyorsa, **Tip 1 Hata** yapılmış olur ("Masumu asmak").

- **Karar 2: Suçsuz Bulma ( $H_0$ 'ı Reddedememe)**

- **Durum:** Kanıtlar, sanığın suçlu olduğunu "makul bir şüphenin ötesinde" kanıtlamak için yeterince güçlü değildir ( $p \geq \alpha$ ).
- **Jüri Kararı:** "Sanığın suçlu olduğunu dair yeterli kanıt bulamadık. Bu yüzden masumiyet karunesini curlyemiyoruz. Sanık suçsuzdur." (Dikkat: "Masum olduğunu kanıtladık" demezler, sadece "suçu olduğunu kanıtlayamadık" derler).
- **Hata Riski:** Eğer sanık gerçekten suçluysa ama kanıtlar yetersiz kaldığı için serbest bırakıldıysa, **Tip 2 Hata** yapılmış olur ("Suçluyu salivermek").

Bu analoji, hipotez testinin neden " $H_0$ 'ı kabul etmek" yerine " **$H_0$ 'ı reddedememek**" ifadesini kullandığını da çok güzel açıklar. Tíkki mahkemenin birinin masumiyetini kanıtlamak yerine, suçluluğunu kanıtlayamaması gibi.

#### Senaryo:

- **Problem:** Yeni bir ilaçın kan basıncını düşürüp düşürmediğini test etmek istiyoruz.
- **Mevcut Durum:** Hasta grubunun mevcut ortalama kan basıncı 140 mmHg. Bu, bizim referans noktamız veya "değişmesini beklediğimiz" değer.

#### Hipotezlerin Kurulması

##### 1. Sıfır Hipotezi ( $H_0$ ) - "Etki Yok" Varsayımlı

Sıfır hipotezi, her zaman "hiçbir değişiklik olmadı", "etki yok" veya "fark yok" şeklinde kurulur. Bu, bizim curlytmeye çalıştığımız başlangıç varsayımlıdır.

- **Sıfır Hipotezi ( $H_0$ ):** Yeni ilaç, kan basıncını **düşürmez** (veya değiştirmez).
- **Matematiksel İfade:** İlacı kullandıkten sonraki yeni ortalama kan basıncı ( $\mu$ ), başlangıçtaki ortalamaya eşit veya ondan daha büyktür.  $H_0: \mu \geq 140 \text{ mmHg}$

Neden " $\geq$ "? Çünkü ilaçın kan basıncını artması da "düşürmediği" anlamına gelir. Bazı durumlarda bu daha basitte  $H_0: \mu = 140$  olarak ifade edilebilir, çünkü testin

odak noktası "değişiklik yok" durumudur.

##### 2. Alternatif Hipotez ( $H_1$ ) - "Etki Var" İddiası

Alternatif hipotez, araştırmacı olarak bizim **kanıtlamaya çalıştığımız** iddiadır. Bu durumda, ilaçın kan basıncını düşürdügüünü kanıtlamak istiyoruz.

- **Alternatif Hipotez ( $H_1$ ):** Yeni ilaç, kan basıncını **düşürür**.
- **Matematiksel İfade:** İlacı kullandıkten sonraki yeni ortalama kan basıncı ( $\mu$ ), başlangıçtaki ortalamadan **daha düşüktür**.  $H_1: \mu < 140 \text{ mmHg}$

#### Özet Tablo

Hipotez Türü	İfade	Anlamı	Matematiksel Gösterim
<b>Sıfır Hipotezi (<math>H_0</math>)</b>	İlaç kan basıncını düşürmez.	Statüko, etki yok.	$\mu \geq 140$
<b>Alternatif Hipotez (<math>H_1</math>)</b>	İlaç kan basıncını düşürür.	Araştırmacının iddiası, etki var.	$\mu < 140$

#### Test Süreci Nasıl İşler?

1. Bir grup hastaya yeni ilaçı veririz.
2. Tedavi sonrası bu grubun ortalama kan basıncını (örneğin,  $132 \text{ mmHg}$ ) ölçeriz.
3. **p-değerini** hesaplarız. Bu, şu soruyu sorar: "Eğer ilaç aslında işe yaramıyor olsaydı ( $H_0$  doğru olsaydı), sadece şans eseri ortalamanın  $132 \text{ mmHg}$  gibi bir değere düşme olasılığı ne olurdu?"
4. Eğer bu olasılık çok düşükse ( $p < 0.05$ ), "Bu kadar büyük bir düşüşün tesadüf olması çok zor. Demek ki ilaç gerçekten işe yarıyor" deriz ve **Sıfır Hipotezi'ni reddederiz**.
5. Eğer olasılık yükseksse ( $p \geq 0.05$ ), "Gördüğümüz bu düşüş şans eseri olabilir. İlacın işe yaradığına dair yeterli kanıt yok" deriz ve **Sıfır Hipotezi'ni reddedemeyiz**.

**" $H_0$  reddeilemez yazacaksınız.  $H_0$  kabul edilir yazmayın. " dedi hoca. bu çok önemli**

" $H_0$  reddeilemez yazacaksınız.  $H_0$  kabul edilir yazmayın. " dedi hoca... hemen özetalıyorum.

Bu ifade doğrudur çünkü;

**Bir şeyi %100 ispatlamak imkansızdır, ama bir şeyi makul bir şüphenin ötesinde curlytmek mümkündür.**

- **$H_1$ 'i (etki var) ispatlamaya çalışmak:** Dünyadaki tüm kuğuların beyaz olduğunu kanıtlamaya çalışmak gibidir. Milyonlarca beyaz kuğu görseniz bile, görmediğiniz bir siyah kuğu olabilir. Bu imkansızdır.
- **$H_0$ 'ı (etki yok)廓清meye çalışmak:** "Siyah kuğu yoktur" iddiasını廓清meyeye çalışmak gibidir. Bunun için **tek bir siyah kuğu bulmanız yeterlidir**. Bu, daha kolay ve daha bilimsel bir yaklaşımdır.

Istatistik, "etki yok" ( $H_0$ ) varsayıminin ne kadar mantıksız olduğunu göstererek çalışır. Eğer bu varsayıım çok mantıksız görünürse (p-değeri düşükse), onu reddederiz ve bu da dolaylı olarak "etki var" ( $H_1$ ) iddiasını destekler.

---

## sınavda SQL sorulacak

python kodu --> SQL e çevrilecek ve SQL --> python koduna çevrilecek.

linuxtan da sorulacak.

---