**Kutu-Bıyık Grafiği(Kutu Grafiği, Box Plot)**

**Yüzdelik Dilim:** Veriler küçükten büyüğe sıralanır ve i endeksi hesaplanır.

p = Yüzdelik Dilim

n = Gözlem Sayısı

\* i tam sayı değil ise bir üst sayıya yuvarla.

\* i tam sayı ise i ve (i+1) ortalaması i endeksini verir.

Örnek: Aşağıdaki veriden 85. Yüzdelik dilimi bulunuz.

15 • 20 • 25 • 30 • 35 • 40 • 45 • 50 • 55 • 60 • 65 • 70

11. terim 65 verinin 85. Yüzdelik dilim değeridir.

Örnek: Aşağıdaki veriden 50. Yüzdelik dilimi bulunuz.

15 • 20 • 25 • 30 • 35 • 40 • 42 • 50 • 55 • 60 • 65 • 70

|  |
| --- |
| Bulduğumuz 65 ve 41 değerleri bize sırasıyla verinin %85’inin ve %50’sinin o skor değerinden küçük olduğunu gösterir, ne kadar küçük olduğunu göstermez. |

6. ve 7. terim skorların ortalaması () değeri verinin 50. Yüzdelik dilimi değeridir. Aynı zamanda medyandır.

Dağılımın özetlenmesinde beş sayı özeti kullanılır.

1. **En küçük değer**
2. **1. Kartil(Dağılımın 25. Yüzdelik dilimi)**
3. **Medyan(Dağılımın 50. Yüzdelik dilimi, 2. Kartil)**
4. **3. Kartil(Dağılımın 75. Yüzdelik dilimi)**
5. **En Büyük Değer**

Örnek: Bir öğrencin matematik ve ingilizce sınavlarının skoru 70’dir. Notları sırasıyla 60. Ve 80. Yüzdelik dilime denk geldiği gözlenmiştir.

1. Notların aynı olması yüzdelik dilimlerin de aynı olmasını gerektirmez mi? Neden?

Hayır çünkü yüzdelik dilim farklı dağılımlarda farklı sonuçlar verir. İlgilendiğimiz dağılımdaki göreceli konumunu belirler. Matematik skorları dağılımındaki 70 skoru 60. Yüzdelik dilime geldiğinden matematik skoru dağılımı için o öğrencinin notu tüm skorların dağılımınının %60’ından daha büyüktür. Ama ingilizce notu için 70 notu yüzdelik dilim %80’dir, 70’notu dağılımın %80’ininin üstündedir.

1. Öğrenci hangi sınavda göreceli olarak daha iyi konumdadır? Neden?

İngilizce sınavında göreceli olarak daha iyi konumdadır çünkü dağılımdaki %80 öğrenciden daha iyi bir not almıştır.

Örnek: 80 kişilik bir sınıfta desten sınıfın %75’i kalmıştır. Geçme notu 60’dır. Durumu yorumlayımız.

3. kartil değeri 60’dır. 60 kişi 60’dan düşük not almıştır. Bu da %75’dir.

**Kartiller Arası Genişlik(IQR):**Değişim genişliğinin serinin iki ucunda yer alan aşırı değerlerden etkilenmesi sakıncasını gidermek için kartiller arası fark kullanılır.

**VERİ SETİ:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 30 | 33 | 34 | 50 | 55 | 57 | 60 | 62 | 63 | 63 |
| 64 | 64 | 65 | 65 | 65 | 66 | 66 | 67 | 67 | 67 |
| 67 | 67 | 67 | 68 | 68 | 68 | 68 | 68 | 69 | 69 |
| 69 | 69 | 69 | 69 | 69 | 70 | 70 | 71 | 71 | 71 |
| 72 | 72 | 73 | 73 | 74 | 74 | 74 | 74 | 75 | 75 |
| 75 | 76 | 76 | 77 | 78 | 79 | 79 | 80 | 80 | 81 |
| 82 | 83 | 83 | 85 | 85 | 86 | 86 | 86 | 87 | 88 |
| 88 | 89 | 89 | 90 | 90 | 90 | 90 | 91 | 91 | 91 |
| 91 | 92 | 92 | 92 | 93 | 93 | 94 | 94 | 94 | 95 |
| 95 | 95 | 96 | 96 | 98 | 100 | 111 | 122 | 125 | 126 |

***BEŞ SAYI ÖZETİ* :** 30, 68, 75, 90, 126

***Uç Değer, Aykırı Değer (Outlier):*** diğer gözlemlerden aykırı ve sapan değerdir.

’den küçük değerler uç değer kabul edilir. (34, 33, 30)

’den buyuk değerler uç değer olarak kabul edilir. (125,126)

**Kutu Bıyık Grafiğinin Python Dilinde Görselleştirilmesi**

import pandas as pd

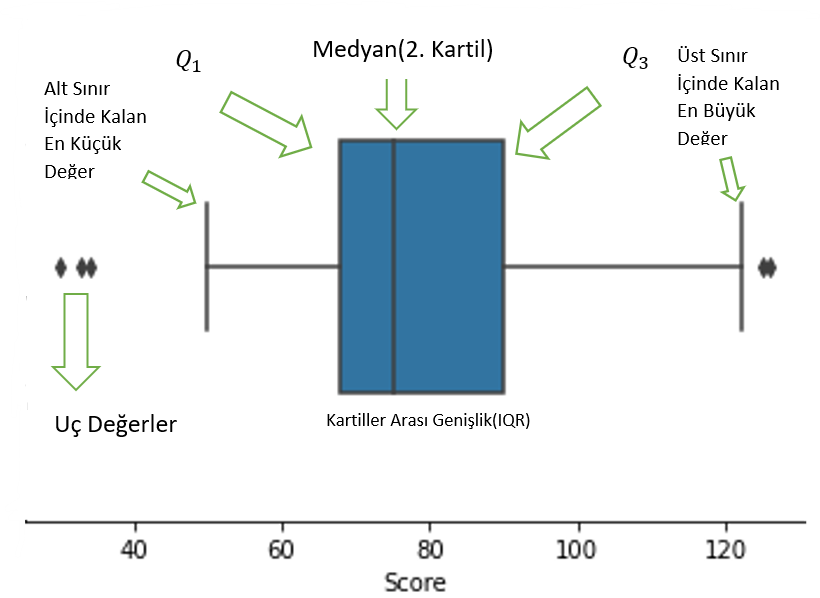
import seaborn as sns

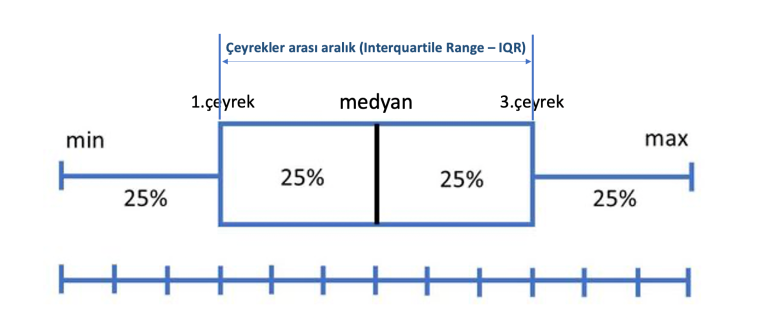
df = pd.read\_csv("ornek.csv")

print(df)

sns.boxplot(x=df["Score"], width=0.5)

**Kutu Bıyık Grafiğinin Görünümü**



****Aşağıdaki resim, tam bir normal dağılan verinin kutu bıyık grafiğidir. Gerçekte bu hiç olmaz.

İşin Özü: Kutu-Bıyık grafiği veri analizinde ve makine öğreniminde uç değerlerin tespit edilmesine ve verideki değişkene ilk bakış atmak, kuş bakışı yapmak için kullanılır. Dağılımın şekli hakkında, Göreceli konumu hakkında bize fikir verir.