

## 9. Ders - Missin Value - Part 3 - MAD

25 Eylül 2025 Perşembe 10:02

# Mean Absolute Deviation

$x_i$	$\bar{x}$	$x_i - \bar{x}$	$ x_i - \bar{x} $
10	16	-6	6
12	16	-4	4
15	16	-1	1
19	16	3	3
24	16	8	8

$$MAD = \frac{\sum |x_i - \bar{x}|}{n}$$

$$\bar{x} = \frac{\sum x_i}{n}$$

MAD, bir dağılımın ortalama mutlak sapmasını (ortalama sapma) ölçer: her gözlemin ortalamadan mutlak farklarının ortalaması. Aykırı değerleri etkilemeyen bir ölçüdür (standart sapma gibi kareli ölçümlerden daha "robust").

```
# Eksik değerleri doldurma örneği
medyan = df['sütun'].median()
mad = (df['sütun'] - medyan).abs().median() # MAD hesaplama

# Medyan ± MAD kullanarak doldurma yapabilirsiniz
```

MAD, kayıp değer doldurma stratejimizi aykırı değerlere karşı korur.

Örnek üzerinden ilerleyelim. Gelir verileri :

1000, 1200, 1500, NaN, 1800, 2000, NaN, 50000

Olan bir veri setinde eksik 2 adet verimiz var. 50.000 değeri ise outlier değerimiz.

Eksik verileri doldururken MAD ve Medyan hesaplayarak bu aralıkta rastgele veri ataması gönül rahatlığıyla yapabiliriz :)

```
import pandas as pd
import numpy as np

# Örnek veri
data = {'gelir': [1000, 1200, 1500, np.nan, 1800, 2000, np.nan, 50000]} # 50000 aykırı değer
df = pd.DataFrame(data)

# MAD ve Medyan hesapla
medyan = df['gelir'].median() # 1650
mad = (df['gelir'].dropna() - medyan).abs().median() # 300

print(f"Medyan: {medyan}, MAD: {mad}")
# Medyan: 1650, MAD: 300
```

Alt ve üst sınırı belirlerken aykırı değer eşiği oluştururuz:

```
# MAD ile aykırı değer eşiği
lower_bound = medyan - 3 * mad # 1650 - 900 = 750
upper_bound = medyan + 3 * mad # 1650 + 900 = 2550

# 50000 değeri 2550'den büyük → Aykırı değer!
```

Şimdi eksik verileri dolduralım :

```
# Eksik değerleri MAD tabanlı doldurma
def mad_based_imputation(column):
    medyan = column.median()
    mad = (column.dropna() - medyan).abs().median()

    # Eksik değerleri medyan ± MAD aralığında rastgele doldur
    missing_mask = column.isnull()
    random_values = np.random.normal(medyan, mad, size=missing_mask.sum())
    column_filled = column.copy()
    column_filled[missing_mask] = random_values

    return column_filled

df['gelir_doldurulmus'] = mad_based_imputation(df['gelir'])
```

Standart Sapmaya göre daha güvenilirdir, özellikle veri setinde aykırı değerler varsa.