



T.C.

Hacettepe Üniversitesi

Fen Fakültesi

İstatistik Bölümü

İST377 PARAMETRİK OLMAYAN İSTATİSTİKSEL YÖNTEMLER DÖNEM ÖDEVİ

**ÖDEV KONUSU:**

Parametrik Olmayan Testlerin Veriler Üzerinde İncelenmesi

**Hazırlayanlar:**

Yasin Çandır 2210329055

**Öğretim Üyeleri**

Doç. Dr. Semra Türkan

Arş. Gör. Dr. Mustafa Murat ARAT

## İçindekiler

Bağımsız Veri Seti .....	3
Tanımlayıcı istatistikler .....	3
Grafikler .....	4
Normallik Varsayımı.....	6
Bağımlı Veri Seti.....	7
Tanımlayıcı istatistikler .....	7
Grafikler .....	8
Normallik Varsayımı.....	10
Trend Veri Seti .....	11
Tanımlayıcı istatistikler .....	11
Grafik .....	11
Normallik Varsayımı.....	11
Testler .....	12
Tek Örneklem Konum Testleri .....	12
İşaret Testi .....	12
Wilcoxon İşaret Sıra Sayıları Testi.....	13
Bağımsız İki Örneklem Konum Testi .....	13
Mann-Whitney U Testi .....	13
Bağımlı İki Örneklem Konum Testi .....	14
Wilcoxon İşaret Sıra Sayıları Testi.....	14
Bağımsız K Örneklem Konum Testi .....	16
Kruskal-Wallis Testi.....	16
Dunn's Post-Hoc Test.....	17
Bağımlı K Örneklem Konum Testi .....	17
Friedman Testi .....	17
Eğilim testi .....	18
Kaynakça.....	19

## Bağımsız Veri Seti

	A	B	C
0	2.0	17.0	29.0
1	12.0	15.0	3.0
2	5.0	3.0	25.0
3	4.0	19.0	28.0
4	26.0	5.0	11.0
5	8.0	14.0	7.0
6	17.0	5.0	5.0
7	4.0	6.0	25.0
8	25.0	19.0	32.0
9	6.0	4.0	24.0
10	21.0	9.0	36.0
11	6.0	7.0	20.0

Her birinde 12 bireyin yer aldığı üç farklı diyet uygulaması sonucunda kişilerin kilolarında meydana gelen azalmalar sol tarafta bulunmuştur

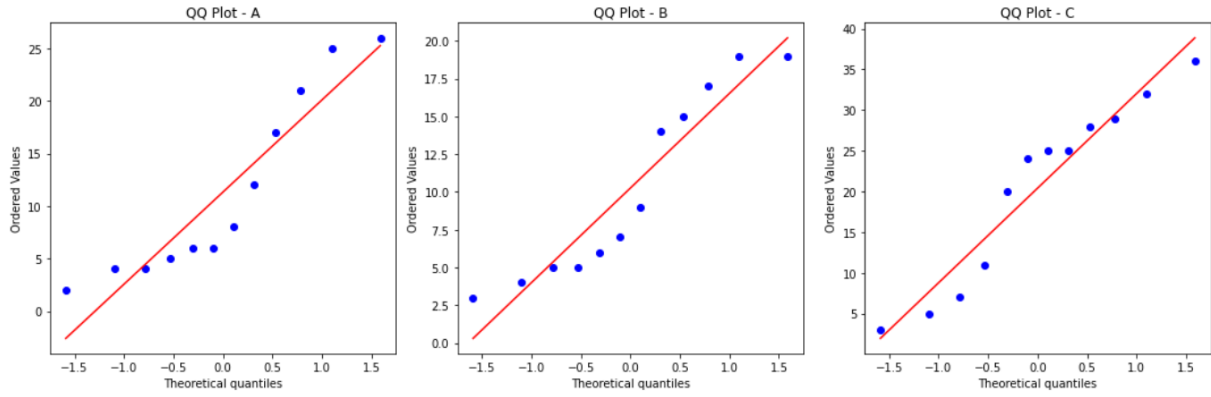
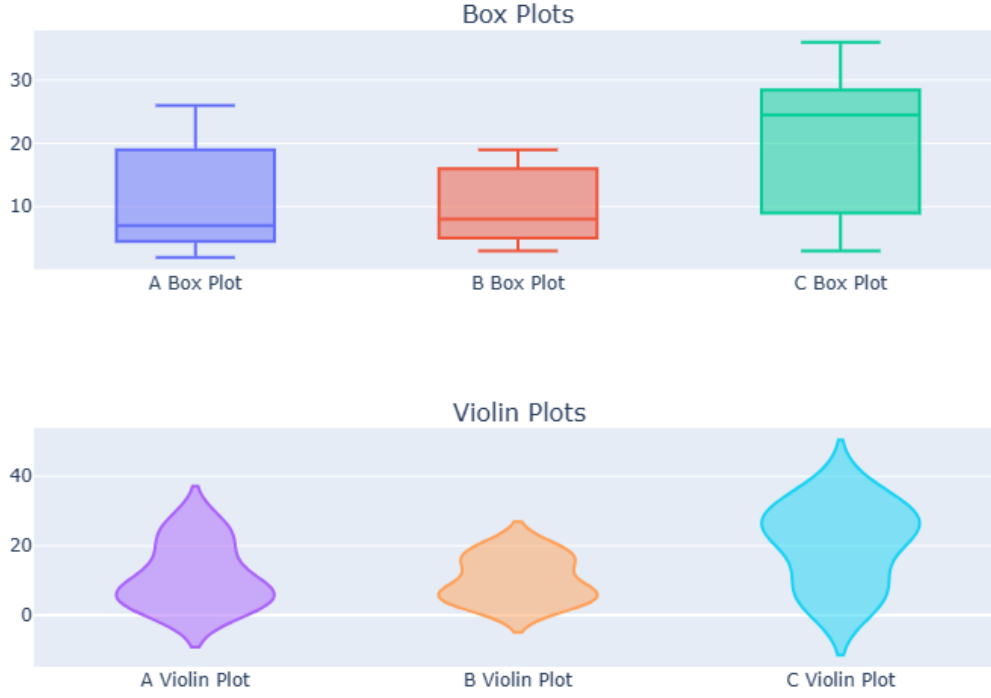
## Tanımlayıcı İstatistikler

	A	B	C
<b>Count</b>	12	12	12
<b>Mean</b>	11.3333	10.25	20.4166
<b>Std Dev</b>	8.6899	6.1218	11.1718
<b>Min</b>	2	3	3
<b>25%</b>	4.75	5	10
<b>50%</b>	7	8	24.5
<b>75%</b>	18	15.5	28.25
<b>Max</b>	26	19	36

- Diyet A'nın ortalaması diğerlerine göre daha yüksektir.
- Diyet B'nin standart sapması diğerlerine göre daha düşüktür, bu dağılımın daha sıkı olduğunu gösterebilir.
- Diyet C'nin minimum değeri diğer değişkenlere göre daha yüksektir. Bu, Diyet C'nin genel olarak daha yüksek değerlere sahip olduğunu gösterebilir.

## Grafikler

### Box Plots and Violin Plots



İlk grafik (boxplot), veri setinin genel dağılımını ve potansiyel aykırı değerleri gösterir. Eğer aykırı değerler varsa, bu grafik onları belirgin şekilde gösterir.

İkinci grafik (violinplot), veri setinin dağılımının genel formunu daha detaylı bir şekilde sunar. Yoğunluk bilgisini içerir ve veri setinin hangi bölgelerde yoğun olduğunu daha net gösterir.

Üçüncü grafik (QQ Plot), veri setinin normal dağılıma ne kadar uygun olduğunu gösterir. Eğer veri normal dağılıma uygunsa, noktalar çizgi üzerinde düzgün bir şekilde dağılmış olacaktır.

Bu bilgiler doğrultusunda ,

Veride aykırı değer olmadığı gözlemlenmektedir.

A ve C diyetinin normalliği sağlamadığını düşünülebilir ancak emin olmak için normallik testlerine başvurulmalıdır.

## Normallik Varsayımı

Veri büyüklüğü 30'dan küçük olduğundan Shapiro-Wilk testi kullanılmıştır.

Sonuçlar aşağıdaki gibidir.

*Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.*

*Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.*

Shapiro-Wilk test for column 'A':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.0409. Null hipotezi reddedilir. Veri seti normal olarak dağılmamıştır.

Shapiro-Wilk test for column 'B':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.0676. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır.

Shapiro-Wilk test for column 'C':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.2308. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır.

A diyetinde %95 güven aralığında P-value:  $0.0409 < \alpha = 0.05$  olduğundan  $H_0$  reddedilir. Veri seti normal olarak dağılmamıştır. Bundan dolayı parametrik olmayan testler tercih edilecektir.

## Bağımlı Veri Seti

Bir pazarlama araştırması için yedi mağaza seçilmiştir. Her mağazada beş farklı markanın yeni tip el losyonu yan yana yer alıyor. Hafta sonunda her marka için satılan losyon şişesi sayısı aşağıdaki sonuçlarla tablolştırılır.

	A	B	C	D	E
1	5	4	7	10	12
2	1	3	1	0	2
3	16	12	22	22	35
4	5	4	3	5	4
5	10	9	7	13	10
6	19	18	28	37	58
7	10	7	6	8	7

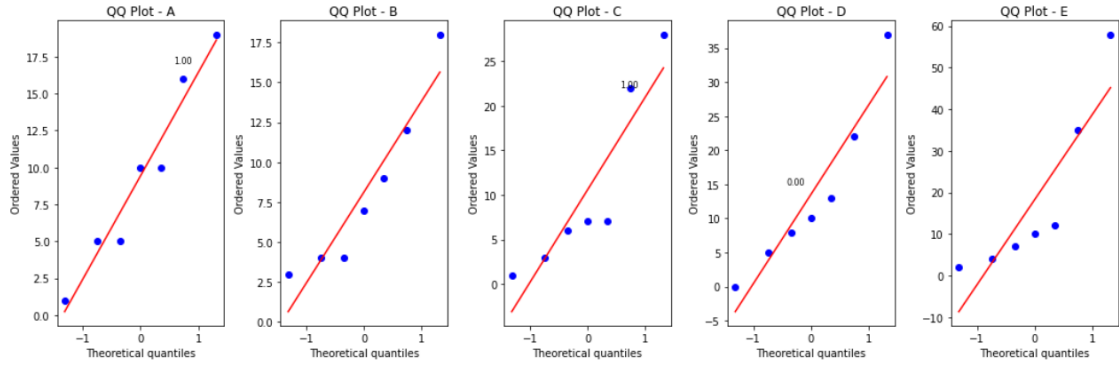
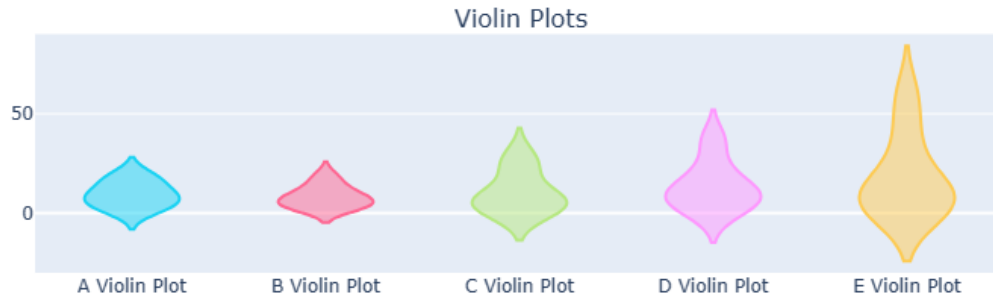
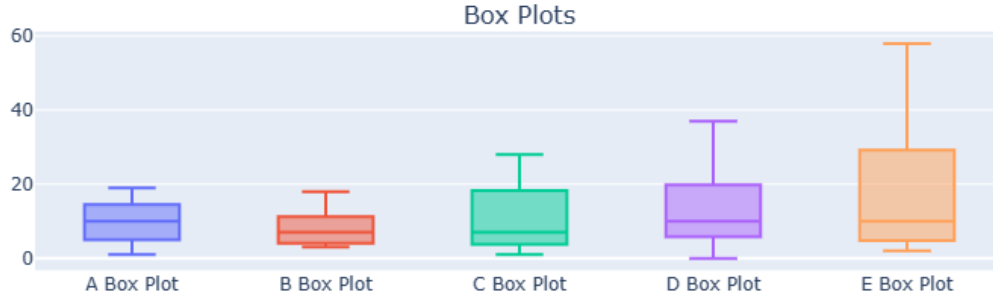
## Tanımlayıcı istatistikler

	A	B	C	D	E
count	7.000000	7.000000	7.000000	7.000000	7.000000
mean	9.428571	8.142857	10.571429	13.571429	18.285714
std	6.399405	5.398412	10.244627	12.394315	20.661674
min	1.000000	3.000000	1.000000	0.000000	2.000000
25%	5.000000	4.000000	4.500000	6.500000	5.500000
50%	10.000000	7.000000	7.000000	10.000000	10.000000
75%	13.000000	10.500000	14.500000	17.500000	23.500000
max	19.000000	18.000000	28.000000	37.000000	58.000000

- Marka A'nın medyan değeri 10'dur ve ortalaması ile benzerdir, bu durum normal bir dağılımı işaret edebilir.
- Marka C'nin medyan değeri 7'dir, ancak ortalaması daha yüksektir. Bu durum, C'nin sağa çarpık bir dağılıma sahip olabileceğini gösterebilir.
- Marka E'nin standart sapması diğer markalara göre belirgin şekilde daha yüksektir. Bu, Marka E'nin değerlerinin genellikle birbirinden daha uzak olduğunu ve daha geniş bir dağılıma sahip olduğunu gösterir.

## Grafikler

### Box Plots and Violin Plots





İlk grafik (boxplot), veri setinin genel dağılımını ve potansiyel aykırı değerleri gösterir. Eğer aykırı değerler varsa, bu grafik onları belirgin şekilde gösterir.

İkinci grafik (violinplot), veri setinin dağılımının genel formunu daha detaylı bir şekilde sunar. Yoğunluk bilgisini içerir ve veri setinin hangi bölgelerde yoğun olduğunu daha net gösterir.

Üçüncü grafik (QQ Plot), veri setinin normal dağılıma ne kadar uygun olduğunu gösterir. Eğer veri normal dağılıma uygunsa, noktalar çizgi üzerinde düzgün bir şekilde dağılmış olacaktır.

Bu bilgiler doğrultusunda ,

Veride aykırı değer olmadığı gözlemlenmektedir.

E marka da satılan losyon şişesi sayısının normalliği sağlamadığını düşünülebilir ancak emin olmak için normallik testlerine başvurulmalıdır.

## Normallik Varsayımı

*Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.*

*Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.*

Shapiro-Wilk test for column 'A':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.7030. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'B':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.2758. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'C':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.0641. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'D':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.3748. Null hipotezi reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'E':

Null Hipotezi ( $H_0$ ): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi ( $H_1$ ): Veri seti normal olarak dağılmamıştır.

Sonuç: p-value = 0.0328. Null hipotezi reddedilir. Veri seti normal olarak dağılmamıştır.

Marka E normal olarak dağılmamıştır.  
Bundan dolayı parametrik olmayan testler tercih edilecektir.

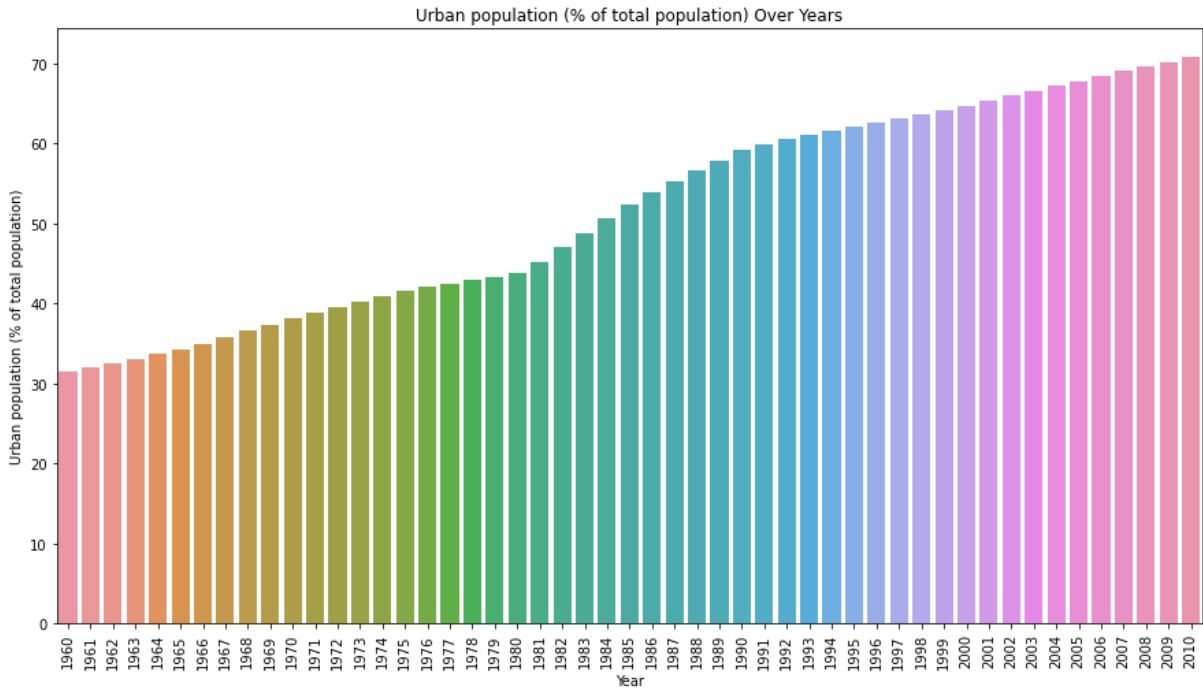
## Trend Veri Seti

Sanayinin ve iş gücünün kentlerde daha yoğun olması nedeniyle, 1960-2010 yılları arasında Türkiye'deki kentleşme oranının arttığı tahmin edilmektedir. Bu doğrultuda, söz konusu dönemdeki kentleşme oranı verileri üzerinde istatistiksel bir inceleme gerçekleştirilecektir.

## Tanımlayıcı istatistikler

	Urban population (% of total population)	Year
Count	51	51
Mean	51.530059	1985
Std Dev	12.98	14.8666069
Min	31.515	1960
25%	39.929	1972.5
50%	52.448	1985
75%	63.441	1997.5
Max	70.825	2010

## Grafik



## Normallik Varsayımı

Test İstatistiği: 1.0000

P Değeri: 0.0000

Örnek veri normal bir dağılıma sahip değildir (H0 reddedilir)

Veri seti normal olarak dağılmamıştır. Bundan dolayı parametrik olmayan testler tercih edilecektir.

## Testler

### Tek Örneklem Konum Testleri

$$H_0: \theta_A = 20$$

$$H_1: \theta_A > 20$$

A diyetinin ortancasının 20'den büyük olup olmadığı araştırılmak istenmektedir

```
In [14]: 1 print("df_independent : ",df_independent["A"])
2 alleged = 20
3 nPlus = len([x for x in df_independent["A"] if x > alleged])
4 nNeg = len([x for x in df_independent["A"] if x < alleged])
5 n = nPlus + nNeg
6
7 print("+ -> ",nPlus)
8 print("- -> ",nNeg)

df_independent : 0      2.0
1      12.0
2       5.0
3       4.0
4      26.0
5       8.0
6      17.0
7       4.0
8      25.0
9       6.0
10     21.0
11       6.0
Name: A, dtype: float64
+ -> 3
- -> 9
```

### İşaret Testi

İşaret Testi

```
In [15]: 1 from scipy.stats import binom_test
2 p_value = stats.binomtest(nPlus, n, p=0.5, alternative='greater')
3 p_value
```

Out[15]: BinomTestResult(k=3, n=12, alternative='greater', statistic=0.25, pvalue=0.980712890625)

```
In [16]: 1 p_value = p_value.pvalue
2 alpha = 0.05
3 if p_value > alpha:
4     decision = f"Elde edilen p değeri {alpha}'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edil
5 else:
6     decision = f"Elde edilen p değeri {alpha}'e eşit veya daha küçüktür. Bu, Ortancanın 20'den büyük olarak kabul etmek için
7
8 print(decision)
9 print(f"İşaret Test Sonucu: p-value = {p_value}")
```

Elde edilen p değeri 0.05'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edilebilir.

İşaret Test Sonucu: p-value = 0.980712890625

- BinomTestResult(k=3, n=12, alternative='greater', statistic=0.25, pvalue=0.980712890625)
- Elde edilen p değeri 0.05'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edilebilir.
- İşaret Test Sonucu: p-value = 0.980712890625

## Wilcoxon İşaret Sıra Sayıları Testi

Wilcoxon İşaret Sıra Sayıları Testi

```
In [17]: 1 from scipy.stats import wilcoxon
2 result = wilcoxon([x - alleged for x in df_independent["C"]], alternative='greater', method='exact', correction=True)
3 result
```

Out[17]: WilcoxonResult(statistic=32.5, pvalue=0.5354585807761036)

```
In [18]: 1 p_value = result.pvalue
2 alpha = 0.05
3 if p_value > alpha:
4     decision = f"Elde edilen p değeri {alpha}'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edil
5 else:
6     decision = f"Elde edilen p değeri {alpha}'e eşit veya daha küçüktür. Bu, Ortancanın 20'den büyük olarak kabul etmek için
7
8 print(decision)
9 print(f"Wilcoxon Test Sonucu: p-value = {p_value}")
```

Elde edilen p değeri 0.05'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edilebilir.

Wilcoxon Test Sonucu: p-value = 0.5354585807761036

- WilcoxonResult(statistic=32.5, pvalue=0.5354585807761036)
- Elde edilen p değeri 0.05'ten büyüktür. Ortancanın 20'den büyük olmadığı %95 güven düzeyi ile ifade edilebilir.
- Wilcoxon Test Sonucu: p-value = 0.5354585807761036

## Bağımsız İki Örneklem Konum Testi

$$H_0: \theta_A - \theta_C = 0$$

$$H_0: \theta_A - \theta_C < 0$$

C diyetinin ortancasının A diyetinin ortancasından büyük olup olmadığı öğrenilmek istenmektedir.

## Mann-Whitney U Testi

```
In [19]: 1 from scipy import stats
2 result = stats.mannwhitneyu(df_independent["A"], df_independent["C"], alternative='less')
3
4 U1 = result.statistic
5 U2 = len(df_independent["A"]) * len(df_independent["C"]) - U1
6
7 print(f"Test İstatistiği: {result.statistic}")
8 print(f"P-değeri: {result.pvalue}")
9 print(f"U1 Değeri: {U1}")
10 print(f"U2 Değeri: {U2}")
```

Test İstatistiği: 38.5  
P-değeri: 0.02818515194903695  
U1 Değeri: 38.5  
U2 Değeri: 105.5

```
In [20]: 1 p_value = result.pvalue
2 alpha = 0.05
3
4 if p_value > alpha:
5     decision = f"Elde edilen p değeri {alpha}'ten büyüktür. İki grup arasında istatistiksel olarak anlamlı bir farklılık bul
6 else:
7     decision = f"Elde edilen p değeri {alpha}'e eşit veya daha küçüktür. İki grup arasında istatistiksel olarak anlamlı bir
8
9 print(f"Mann-Whitney U Test Sonucu: p-value = {p_value}\n")
10 print(decision)
```

Mann-Whitney U Test Sonucu: p-value = 0.02818515194903695

Elde edilen p değeri 0.05'e eşit veya daha küçüktür. İki grup arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.  
A diyeti ile C diyetinin ortancaları arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.

Test İstatistiği: 38.5  
P-değeri: 0.02818515194903695  
U1 Değeri: 38.5  
U2 Değeri: 105.5

Mann-Whitney U Test Sonucu: p-value = 0.02818515194903695

Elde edilen p değeri 0.05'e eşit veya daha küçüktür. İki grup arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.

A diyeti ile C diyetinin ortancaları arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.

## Bağımlı İki Örneklem Konum Testi

$$H_0: \theta_A - \theta_B = 0$$

$$H_0: \theta_A - \theta_B > 0$$

A markada satılan losyon şişesi sayısının B markada satılan losyon şişesi sayısından fazla olup olmadığı araştırılmak istenmektedir.

## Wilcoxon İşaret Sıra Sayıları Testi

Wilcoxon İşaret Sıra Sayıları Testi (Bağımlı veri)

```
In [21]: 1 from scipy.stats import wilcoxon
2 wilcoxon_result = stats.wilcoxon(df_dependent['A'], df_dependent['B'], alternative = 'greater', method='exact', correction=
3 wilcoxon_result
```

Out[21]: WilcoxonResult(statistic=23.0, pvalue=0.078125)

```
In [22]: 1 wilcoxon_p_value = wilcoxon_result.pvalue
2 alpha_wilcoxon = 0.05
3
4 if wilcoxon_p_value > alpha_wilcoxon:
5     wilcoxon_decision = f"Elde edilen p değeri {alpha_wilcoxon}'ten büyüktür. İki grup arasında istatistiksel olarak anlamlı
6 else:
7     wilcoxon_decision = f"Elde edilen p değeri {alpha_wilcoxon}'e eşit veya daha küçüktür. İki grup arasında istatistiksel o
8
9 print(f"Wilcoxon Test Sonucu: p-value = {wilcoxon_p_value}\n")
10 print(wilcoxon_decision)
```

Wilcoxon Test Sonucu: p-value = 0.078125

Elde edilen p değeri 0.05'ten büyüktür. İki grup arasında istatistiksel olarak anlamlı bir farklılık bulunmamaktadır. A markada satılan losyon şişesi sayısı B markada satılan losyon şişesi sayısından fazladır denilemez.

```

In [24]: 1 group1 = np.array([5.00, 1.00, 16.00, 5.00, 10.00, 19.00, 10.00])
2 group2 = np.array([4.00, 3.00, 12.00, 4.00, 9.00, 18.00, 7.00])
3
4 print("Gruplar farkı", group1 - group2)
5
6 absolute_diff = np.abs(group1 - group2)
7 print("absolute_diff", absolute_diff)
8
9 sorted_diff_indices = np.argsort(absolute_diff)
10 sorted_diff = absolute_diff[sorted_diff_indices]
11 sorted_diff = sorted_diff [sorted_diff>0]
12
13 unique_diff, counts = np.unique(sorted_diff, return_counts=True)
14
15 new_ranks = []
16 current_rank = 1
17
18 for i, count in enumerate(counts):
19     new_ranks.extend([current_rank + (count - 1) / 2] * count)
20     current_rank += count
21
22 print(f"Sıralı Mutlak Farklar: {sorted_diff}")
23 print(f"Yeni Rank Dizisi : {new_ranks}")
24
25 positive_diff_values = absolute_diff[group1 - group2 > 0]
26 print(positive_diff_values)
27
28 toplam = 0
29 for value in positive_diff_values:
30     corresponding_rank = new_ranks[np.where(sorted_diff == value)[0][0]]
31     print(f"Pozitif Değer: {value}, Karşılık gelen new_rank: {corresponding_rank}")
32     toplam += corresponding_rank
33
34 print(f"Wilcoxon Test İstatistiği (W) Değeri: {toplam}")

```

```

Gruplar farkı [ 1. -2.  4.  1.  1.  1.  3.]
absolute_diff [1.  2.  4.  1.  1.  1.  3.]
Sıralı Mutlak Farklar: [1.  1.  1.  1.  2.  3.  4.]
Yeni Rank Dizisi : [2.5, 2.5, 2.5, 2.5, 5.0, 6.0, 7.0]
[1.  4.  1.  1.  1.  1.  3.]
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 4.0, Karşılık gelen new_rank: 7.0
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 3.0, Karşılık gelen new_rank: 6.0
Wilcoxon Test İstatistiği (W) Değeri: 23.0

```

WilcoxonResult(statistic=23.0, pvalue=0.078125)

```

Gruplar farkı [ 1. -2.  4.  1.  1.  1.  3.]
absolute_diff [1.  2.  4.  1.  1.  1.  3.]
Sıralı Mutlak Farklar: [1.  1.  1.  1.  2.  3.  4.]
Yeni Rank Dizisi : [2.5, 2.5, 2.5, 2.5, 5.0, 6.0, 7.0]
[1.  4.  1.  1.  1.  1.  3.]
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 4.0, Karşılık gelen new_rank: 7.0
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 1.0, Karşılık gelen new_rank: 2.5
Pozitif Değer: 3.0, Karşılık gelen new_rank: 6.0
Wilcoxon Test İstatistiği (W) Değeri: 23.0

```

Elde edilen p değeri 0.05'ten büyüktür. İki grup arasında istatistiksel olarak anlamlı bir farklılık bulunmamaktadır.

A markada satılan losyon şişesi sayısı B markada satılan losyon şişesi sayısından fazladır denilemez

## Bağımsız K Örneklem Konum Testi

$$H_0: \theta_1 = \theta_2 = \theta_3$$

$$H_1: \theta_i \neq \theta_j \text{ (En az bir çift farklılık göstermektedir)}$$

A, B ve C diyetleri arasında istatistiksel olarak anlamlı bir farklılık olup olmadığı araştırılmak istenmektedir.

## Kruskal-Wallis Testi

Bağımsız K Örneklem Konum Testi

 $H_0: \theta_1 = \theta_2 = \theta_3$  $H_1: \theta_i \neq \theta_j$  (En az bir çift farklılık göstermektedir.)

```
In [24]: 1 from scipy.stats import kruskal
2 result = stats.kruskal(df_independent["A"], df_independent["B"], df_independent["C"])
3 result
```

```
Out[24]: KruskalResult(statistic=6.138031594352627, pvalue=0.04646686517954499)
```

```
In [25]: 1 kruskal_p_value = result.pvalue
2 alpha_kruskal = 0.05
3
4 if kruskal_p_value > alpha_kruskal:
5     kruskal_decision = f"Elde edilen p değeri {alpha_kruskal}'ten büyüktür. Gruplar arasında istatistiksel olarak anlamlı bir
6 else:
7     kruskal_decision = f"Elde edilen p değeri {alpha_kruskal}'e eşit veya daha küçüktür. Gruplar arasında istatistiksel olarak anlamlı bir
8
9 print(f"Kruskal-Wallis Test Sonucu: p-value = {kruskal_p_value}\n")
10 print(kruskal_decision)
```

Kruskal-Wallis Test Sonucu: p-value = 0.04646686517954499

Elde edilen p değeri 0.05'e eşit veya daha küçüktür. Gruplar arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır. A, B ve C diyetleri arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.

Gruplar arasındaki farklılığı bulmak için post-hoc testlere başvurulur.

```
In [26]: 1 from scikit_posthocs import posthoc_dunn
2 # https://www.theanalysisfactor.com/dunns-test-post-hoc-test-after-kruskal-wallis/
3
4 kruskal_stat, p_value = kruskal(df_independent['A'], df_independent['B'], df_independent['C'])
5
6 posthoc_dunn_result = posthoc_dunn(df_independent.melt(value_name='value', var_name='groups'), val_col='value', group_col='groups')
7
8 print("Kruskal-Wallis Test Statistic:", kruskal_stat)
9 print("P-value:", p_value)
10 print("\nDunn's Post-Hoc Test Results:")
11 print(posthoc_dunn_result)
```

Kruskal-Wallis Test Statistic: 6.138031594352627  
P-value: 0.04646686517954499

Dunn's Post-Hoc Test Results:

	A	B	C
A	1.000000	0.853725	0.040614
B	0.853725	1.000000	0.025627
C	0.040614	0.025627	1.000000

A-C ve B-C diyetleri arasında %95 güven düzeyinde anlamlı bir fark vardır.

- `KruskalResult(statistic=6.138031594352627, pvalue=0.04646686517954499)`
- Kruskal-Wallis Test Sonucu: p-value = 0.04646686517954499
- Elde edilen p değeri 0.05'e eşit veya daha küçüktür. Gruplar arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.
- A, B ve C diyetleri arasında istatistiksel olarak anlamlı bir farklılık bulunmaktadır.
- Gruplar arasındaki farklılığı bulmak için post-hoc testlere başvurulur.



## Dunn's Post-Hoc Test

Dunn's Post-Hoc Test Results:

	A	B	C
A	1.000000	0.853725	0.040614
B	0.853725	1.000000	0.025627
C	0.040614	0.025627	1.000000

A-C ve B-C diyetleri arasında %95 güven düzeyinde anlamlı bir fark vardır.

## Bağımlı K Örneklem Konum Testi

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5$$

$$H_1: \theta_i \neq \theta_j \text{ (En az bir çift farklılık göstermektedir)}$$

Markaların satılan losyon şişesi bakımından aralarında fark olup olmadığı araştırılmak istenmektedir

## Friedman Testi

Bağımlı K Örneklem Konum Testi

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5$$

$$H_1: \theta_i \neq \theta_j \text{ (En az bir çift farklılık göstermektedir.)}$$

```
In [27]: 1 from scipy.stats import friedmanchisquare
2 result = stats.friedmanchisquare(df_dependent["A"], df_dependent["B"], df_dependent["C"], df_dependent["D"], df_dependent["E"])
3 result
```

```
Out[27]: FriedmanchisquareResult(statistic=8.328358208955212, pvalue=0.08026409058182589)
```

```
In [28]: 1 friedman_p_value = result.pvalue
2 alpha_friedman = 0.05
3
4 if friedman_p_value > alpha_friedman:
5     friedman_decision = f"Elde edilen p değeri {alpha_friedman}'ten büyüktür. Gruplar arasında istatistiksel olarak anlamlı
6 else:
7     friedman_decision = f"Elde edilen p değeri {alpha_friedman}'e eşit veya daha küçüktür. Gruplar arasında istatistiksel ol
8
9 print(f"Friedman Test Sonucu: p-value = {friedman_p_value}\n")
10 print(friedman_decision)
```

```
Friedman Test Sonucu: p-value = 0.08026409058182589
```

```
Elde edilen p değeri 0.05'ten büyüktür. Gruplar arasında istatistiksel olarak anlamlı bir farklılık bulunmamaktadır.
```

- `FriedmanchisquareResult(statistic=8.328358208955212, pvalue=0.08026409058182589)`
- Friedman Test Sonucu: p-value = 0.08026409058182589
- Elde edilen p değeri 0.05'ten büyüktür. Gruplar arasında istatistiksel olarak anlamlı bir farklılık bulunmamaktadır.
- Markaların satılan losyon şişesi bakımından aralarında fark olmadığı %95 güven düzeyinde söylenebilir.

## Eğilim testi

 $H_0$ : Trend yoktur $H_1$ : (Azalan veya artan) trend vardır

1960-2010 yılları arasında Türkiye'deki kentleşme oranı trendinin olup olmadığı araştırılmak istenmektedir.

Eğilim testi

Ho : Trend yoktur

Hs : Trend vardır / Artan trend vardır / Azalan trend vardır

```
In [30]: 1 import pandas as pd
2 import pymannkendall as mk
3 ## https://pypi.org/project/pymannkendall/
4
5 result = mk.original_test(df_trend["Urban population (% of total population)"], alpha=0.05)
6 print(result)

Mann_Kendall_Test(trend='increasing', h=True, p=0.0, z=10.347696957557925, Tau=1.0, s=1275.0, var_s=15158.333333333334, slope=
0.8374545454545451, intercept=31.511636363636374)

In [31]: 1 trend_type = result.trend
2
3 if trend_type == 'increasing':
4     kendall_decision = f"Mann-Kendall Test Sonucu: Artan bir trend bulunmaktadır."
5 elif trend_type == 'decreasing':
6     kendall_decision = f"Mann-Kendall Test Sonucu: Azalan bir trend bulunmaktadır."
7 else:
8     kendall_decision = f"Mann-Kendall Test Sonucu: Trend bulunmamaktadır."
9
10 kendall_p_value = result.p
11 alpha_kendall = 0.05
12
13 if kendall_p_value > alpha_kendall:
14     kendall_decision += f"\nElde edilen p değeri {alpha_kendall}'ten büyüktür. İstatistiksel olarak anlamlı bir ilişki bulun
15 else:
16     kendall_decision += f"\nElde edilen p değeri {alpha_kendall}'e eşit veya daha küçüktür. İstatistiksel olarak anlamlı bir
17
18 print(kendall_decision)

Mann-Kendall Test Sonucu: Artan bir trend bulunmaktadır.
Elde edilen p değeri 0.05'e eşit veya daha küçüktür. İstatistiksel olarak anlamlı bir ilişki bulunmaktadır.
```

```
Mann_Kendall_Test(trend='increasing', h=True, p=0.0, z=10.347696957557925, Tau=1.0, s=1275.0
,
var_s=15158.333333333334, slope=0.8374545454545451, intercept=31.511636363636374)
```

Mann-Kendall Test Sonucu: Artan bir trend bulunmaktadır.  
Elde edilen p değeri 0.05'e eşit veya daha küçüktür. İstatistiksel olarak anlamlı bir ilişki  
bulunmaktadır.

## Kaynakça

**Conover, William Jay.** *PRACTICAL NONPARAMETRIC STATISTICS Third Edition*. ISBN: 978-0-471-16068-7.

**Doğan, Prof. Dr. İsmet and Doğan, Doç. Dr. Nurhan.** Parametrik olmayan istatistiksel yöntemler. *Parametrik olmayan istatistiksel yöntemler*. s.l. : Detay Yayıncılık, p. 85.

**NumPy Org.** NumPy: the absolute basics for beginners. [Online]  
[https://numpy.org/doc/stable/user/absolute\\_beginners.html#numpy-the-absolute-basics-for-beginners](https://numpy.org/doc/stable/user/absolute_beginners.html#numpy-the-absolute-basics-for-beginners).

**Our World in Data.** Share-of-Population Urban. [Online] <https://ourworldindata.org/grapher/share-of-population-urban?tab=chart&time=earliest..2000&country=~TUR>, 2024.

**pandas.** pandas.DataFrame. [Çevrimiçi] <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html>.

**PyPI.** pymanndkendall 1.4.3. [Çevrimiçi] <https://pypi.org/project/pymanndkendall/>.

**SciPy.** SciPy User Guide. [Çevrimiçi] <https://docs.scipy.org/doc/scipy/tutorial/index.html#user-guide>.

**seaborn.** seaborn.Implot. [Çevrimiçi]  
<https://seaborn.pydata.org/generated/seaborn.Implot.html#seaborn-Implot>.