A red and black logo

Description automatically generated

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](#_Toc155271934)

[Tanımlayıcı istatistikler 3](#_Toc155271935)

[Grafikler 4](#_Toc155271936)

[Normallik Varsayımı 6](#_Toc155271937)

[Bağımlı Veri Seti 7](#_Toc155271938)

[Tanımlayıcı istatistikler 7](#_Toc155271939)

[Grafikler 8](#_Toc155271940)

[Normallik Varsayımı 10](#_Toc155271941)

[Trend Veri Seti 11](#_Toc155271942)

[Tanımlayıcı istatistikler 11](#_Toc155271943)

[Grafik 11](#_Toc155271944)

[Normallik Varsayımı 11](#_Toc155271945)

[Testler 12](#_Toc155271946)

[Tek Örneklem Konum Testleri 12](#_Toc155271947)

[İşaret Testi 12](#_Toc155271948)

[Wilcoxon İşaret Sıra Sayıları Testi 13](#_Toc155271949)

[Bağımsız İki Örneklem Konum Testi 13](#_Toc155271950)

[Mann-Whitney U Testi 13](#_Toc155271951)

[Bağımlı İki Örneklem Konum Testi 14](#_Toc155271952)

[Wilcoxon İşaret Sıra Sayıları Testi 14](#_Toc155271953)

[Bağımsız K Örneklem Konum Testi 16](#_Toc155271954)

[Kruskal-Wallis Testi 16](#_Toc155271955)

[Dunn’s Post-Hoc Test 17](#_Toc155271956)

[Bağımlı K Örneklem Konum Testi 17](#_Toc155271957)

[Friedman Testi 17](#_Toc155271958)

[Eğilim testi 18](#_Toc155271959)

[Kaynakça 19](#_Toc155271960)

# Bağımsız Veri Seti

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

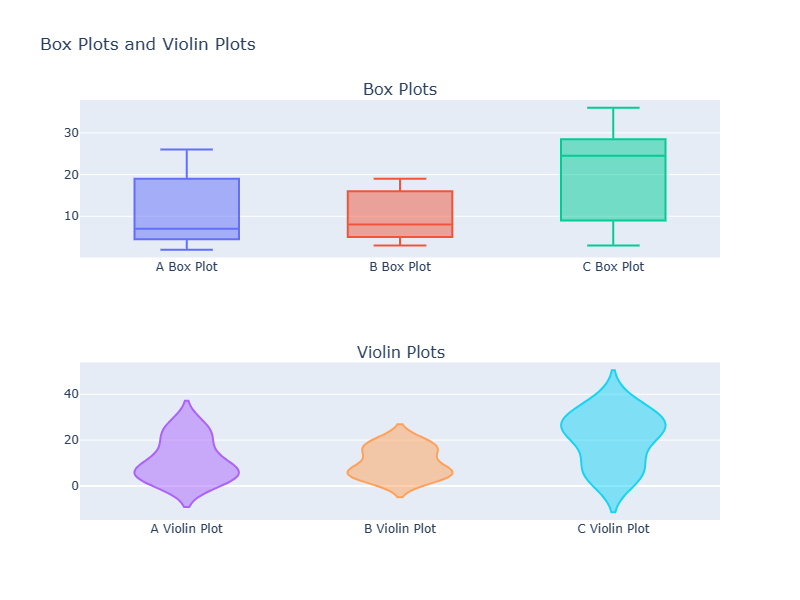
|  |  |  |  |
| --- | --- | --- | --- |
|  | A | B | C |
| 0 | 2.0 | 17.0 | 29.0 |
| 1 | 12.0 | 15.0 | 3.0 |
| 2 | 5.0 | 3.0 | 25.0 |
| 3 | 4.0 | 19.0 | 28.0 |
| 4 | 26.0 | 5.0 | 11.0 |
| 5 | 8.0 | 14.0 | 7.0 |
| 6 | 17.0 | 5.0 | 5.0 |
| 7 | 4.0 | 6.0 | 25.0 |
| 8 | 25.0 | 19.0 | 32.0 |
| 9 | 6.0 | 4.0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

## Tanımlayıcı istatistikler

|  |  |  |  |
| --- | --- | --- | --- |
|  | A | B | C |
| Count | 12 | 12 | 12 |
| Mean | 11.3333 | 10.25 | 20.4166 |
| Std Dev | 8.6899 | 6.1218 | 11.1718 |
| Min | 2 | 3 | 3 |
| 25% | 4.75 | 5 | 10 |
| 50% | 7 | 8 | 24.5 |
| 75% | 18 | 15.5 | 28.25 |
| Max | 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



A diagram of a graph

Description automatically generated

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.

Shapiro-Wilk test for column 'A':

Null Hipotezi (H0): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi (H1): 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 (H0): Veri seti normal olarak

dağılmıştır.

Alternatif Hipotezi (H1): 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 (H0): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi (H1): 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 < α =0.05 olduğundan 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 tablolaştırılır.

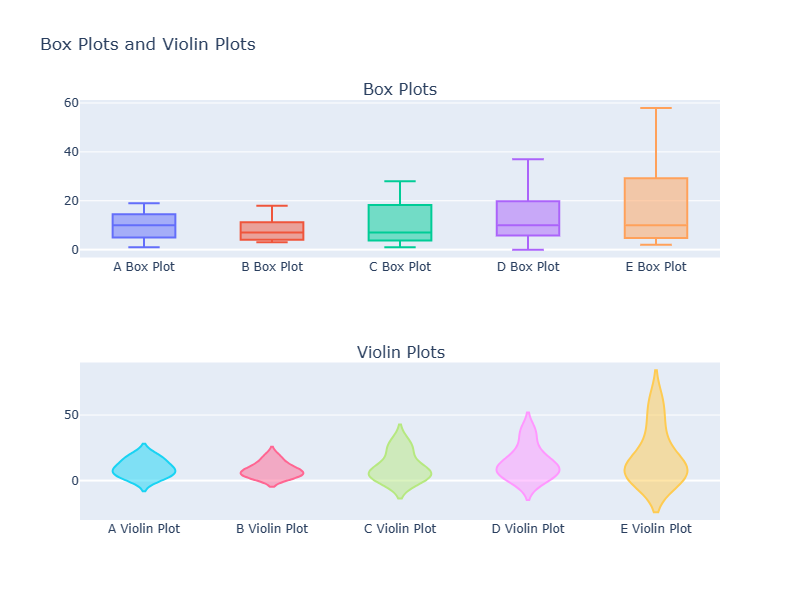
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| 1 | 5 | 4 | 7 | 10 | 12 |
| 2 | 1 | 3 | 1 | 0 | 2 |
| 3 | 16 | 12 | 22 | 22 | 35 |
| 4 | 5 | 4 | 3 | 5 | 4 |
| 5 | 10 | 9 | 7 | 13 | 10 |
| 6 | 19 | 18 | 28 | 37 | 58 |
| 7 | 10 | 7 | 6 | 8 | 7 |

## 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



A diagram of a graph

Description automatically generated

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

Shapiro-Wilk test for column 'B':

Null Hipotezi (H0): Veri seti normal olarak

dağılmıştır.

Alternatif Hipotezi (H1): Veri seti normal

olarak dağılmamıştır.

Sonuç: p-value = 0.2758. Null hipotezi

reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'A':

Null Hipotezi (H0): Veri seti normal olarak

dağılmıştır.

Alternatif Hipotezi (H1): Veri seti normal

olarak dağılmamıştır.

Sonuç: p-value = 0.7030. Null hipotezi

reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'D':

Null Hipotezi (H0): Veri seti normal olarak

dağılmıştır.

Alternatif Hipotezi (H1): Veri seti normal

olarak dağılmamıştır.

Sonuç: p-value = 0.3748. Null hipotezi

reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'C':

Null Hipotezi (H0): Veri seti normal olarak

dağılmıştır.

Alternatif Hipotezi (H1): Veri seti normal

olarak dağılmamıştır.

Sonuç: p-value = 0.0641. Null hipotezi

reddedilemez. Veri seti normal olarak dağılmıştır

Shapiro-Wilk test for column 'E':

Null Hipotezi (H0): Veri seti normal olarak dağılmıştır.

Alternatif Hipotezi (H1): 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

A rainbow colored graph

Description automatically generated

## 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

A diyetinin ortancasının 20’den büyük olup olmadığı araştırılmak istenmektedir

.



### İşaret Testi

A screenshot of a computer program

Description automatically generated

* 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

A screenshot of a computer

Description automatically generated

* 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

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

### Mann-Whitney U Testi

A screenshot of a computer program

Description automatically generated

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

## Bağımlı İki Örneklem Konum Testi

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

### Wilcoxon İşaret Sıra Sayıları Testi

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

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. 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

A, B ve C diyetleri arasında

istatistiksel olarak anlamlı bir

farklılık olup olmadığı araştırılmak

istenmektedir.

### Kruskal-Wallis Testi

A screenshot of a computer

Description automatically generated

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

A B C

A 1.000000 0.853725 0.040614

B 0.853725 1.000000 0.025627

C 0.040614 0.025627 1.000000

## Bağımlı K Örneklem Konum Testi

Markaların satılan losyon şişesi bakımından aralarında fark olup olmadığı araştırılmak istenmektedir

### Friedman Testi

A screenshot of a computer program

Description automatically generated

* 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

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

A screenshot of a computer program

Description automatically generated

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.

**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.** pymannkendall 1.4.3. [Çevrimiçi] https://pypi.org/project/pymannkendall/.

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

**seaborn.** seaborn.lmplot. [Çevrimiçi] https://seaborn.pydata.org/generated/seaborn.lmplot.html#seaborn-lmplot.