Veri analizi 1. Proje ödevi

Merhaba, ödevimi ekim ve kasım ayı akaryakıt fiyatları üzerine yapmak istedim veriler aşağıdadır.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| date | benzin | motorin | yag | ay |
| 1.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 2.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 3.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 4.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 5.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 6.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 7.10.2021 | 7,76 | 7,3 | 8,26 | ekim |
| 8.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 9.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 10.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 11.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 12.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 13.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 14.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 15.10.2021 | 7,76 | 7,53 | 8,63 | ekim |
| 16.10.2021 | 7,92 | 7,96 | 9,03 | ekim |
| 17.10.2021 | 7,92 | 7,96 | 9,03 | ekim |
| 18.10.2021 | 7,92 | 7,96 | 9,03 | ekim |
| 19.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 20.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 21.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 22.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 23.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 24.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 25.10.2021 | 7,92 | 8,19 | 9,27 | ekim |
| 26.10.2021 | 8,37 | 8,19 | 9,27 | ekim |
| 27.10.2021 | 8,65 | 8,19 | 9,27 | ekim |
| 28.10.2021 | 8,4 | 8,19 | 9,27 | ekim |
| 29.10.2021 | 8,41 | 8,19 | 9,27 | ekim |
| 30.10.2021 | 8,41 | 8,19 | 9,27 | ekim |
| 31.10.2021 | 8,41 | 8,19 | 9,27 | ekim |
| 1.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 2.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 3.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 4.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 5.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 6.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 7.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 8.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 9.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 10.11.2021 | 8,41 | 8,19 | 9,27 | kasım |
| 11.11.2021 | 8,09 | 8,19 | 9,27 | kasım |
| 12.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 13.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 14.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 15.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 16.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 17.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 18.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 19.11.2021 | 8,13 | 8,22 | 9,27 | kasım |
| 20.11.2021 | 8,65 | 8,72 | 9,81 | kasım |
| 21.11.2021 | 8,65 | 8,72 | 9,81 | kasım |
| 22.11.2021 | 8,65 | 8,72 | 9,81 | kasım |
| 23.11.2021 | 8,65 | 8,72 | 9,81 | kasım |
| 24.11.2021 | 8,65 | 8,72 | 9,81 | kasım |
| 25.11.2021 | 9,67 | 9,78 | 10,94 | kasım |
| 26.11.2021 | 9,67 | 9,78 | 11,39 | kasım |
| 27.11.2021 | 9,67 | 9,78 | 10,87 | kasım |
| 28.11.2021 | 9,67 | 9,78 | 10,87 | kasım |
| 29.11.2021 | 9,67 | 9,78 | 10,87 | kasım |
| 30.11.2021 | 9,67 | 9,78 | 9,81 | kasım |

İlk olarak temel istatistik ve normallik dağılımı için spss kullandım.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case Processing Summary** | | | | | | |
|  | Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| benzin | 61 | 100,0% | 0 | 0,0% | 61 | 100,0% |
| motorin | 61 | 100,0% | 0 | 0,0% | 61 | 100,0% |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Descriptives** | | | | |
|  | | | Statistic | Std. Error |
| benzin | Mean | | 8,2746 | ,07056 |
| 95% Confidence Interval for Mean | Lower Bound | 8,1334 |  |
| Upper Bound | 8,4157 |  |
| 5% Trimmed Mean | | 8,2257 |  |
| Median | | 8,1300 |  |
| Variance | | ,304 |  |
| Std. Deviation | | ,55110 |  |
| Minimum | | 7,76 |  |
| Maximum | | 9,67 |  |
| Range | | 1,91 |  |
| Interquartile Range | | ,57 |  |
| Skewness | | 1,455 | ,306 |
| Kurtosis | | 1,654 | ,604 |
| motorin | Mean | | 8,1938 | ,08391 |
| 95% Confidence Interval for Mean | Lower Bound | 8,0259 |  |
| Upper Bound | 8,3616 |  |
| 5% Trimmed Mean | | 8,1553 |  |
| Median | | 8,1900 |  |
| Variance | | ,429 |  |
| Std. Deviation | | ,65533 |  |
| Minimum | | 7,30 |  |
| Maximum | | 9,78 |  |
| Range | | 2,48 |  |
| Interquartile Range | | ,48 |  |
| Skewness | | 1,059 | ,306 |
| Kurtosis | | 1,278 | ,604 |

Skewness ve kurtosis değerleri +2 ve -2 arasında kalıyorsa normal kabul edebiliriz.

Bunun içinde kaynak: George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson gösterebiliriz.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| benzin | ,206 | 61 | ,000 | ,791 | 61 | ,000 |
| motorin | ,304 | 61 | ,000 | ,812 | 61 | ,000 |
| a. Lilliefors Significance Correction | | | | | | |

“Normal dağılım testi için ortalama değeri (mean) ve median (medyan) değerinin birbirine yakın sonuçlar olması, veri setinin normal dağılıma sahip olduğunu göstermektedir.”

Bu ifadeye göre baktığım zaman ;

Benzin mean =8,2746 ve median=8,1300 arasında biraz fark var yorum yapmak için diğer yöntemlere de bakalım

Motorin mean=8,1938 ve median=8,1900 fark çok yakın ama diğer yöntemleri de kullanalım.

Skewness (Çarpıklık) ve Kurtosis (Basıklık) değeri normal dağılım testi için bakılması gereken iki değer. Skewness ve Kurtosis değerlerinin normal dağılımı göstermesi için 0 ya da 0’a en yakın değerlere sahip olması gerekmektedir.”

Benzin skewness=1,455 Kurtosis=1,654 0 yakın değerler olduğunu görüyoruz

Motorin skawness= 1,059 Kurtosis= 1,278 ama yine de diğer yöntemleri ve testleri ………………………………………………………….. yapmadan bir şey söylemiyoruz

………………………………………………………………………………………………………………………………

Normal dağılım için z testi uygularsak

Skewness değerinin standart error bölünmesiyle elde ettiğimiz değerin

-1,96<x>1,96 arasında olup olmadığına bakacağız.

Benzin için 1,455 /,07056= 20,6207 aralıktadır

Motorin için : 1,059 /,306 = 3,4607 aralıktadır

………………………………………………………………………..

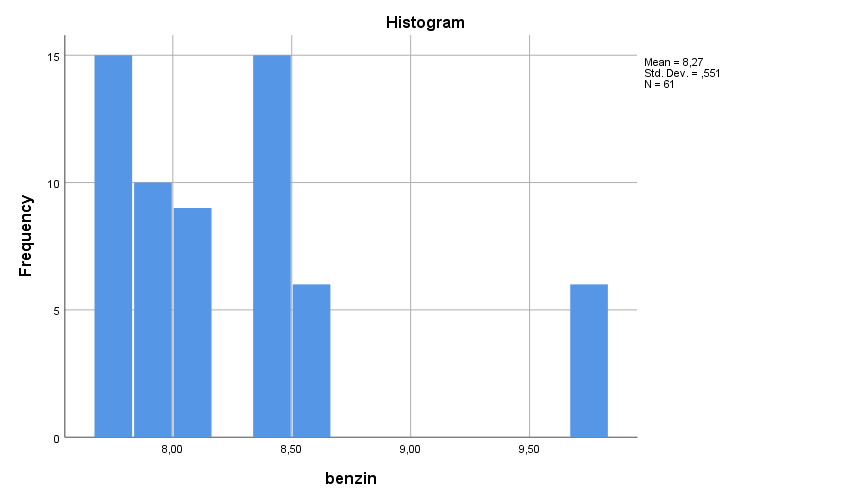
Kolmogorov-Smirnov normallik testi :

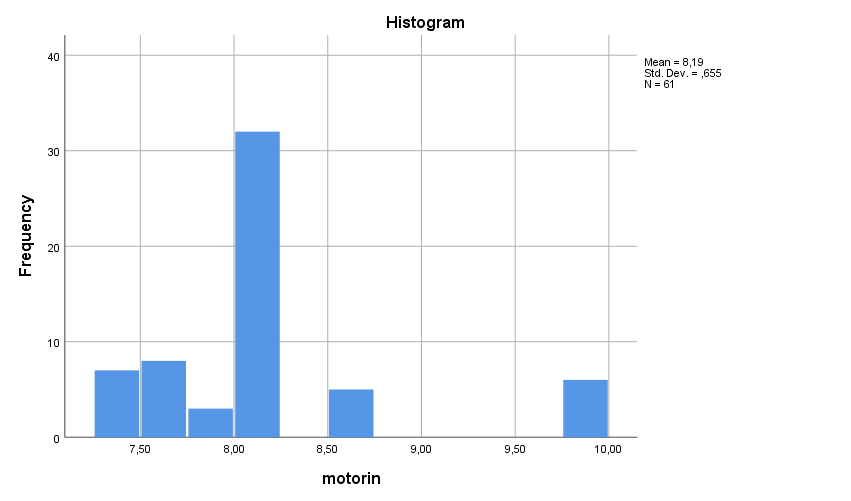
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | |
|  | Kolmogorov-Smirnova | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |
| benzin | ,206 | 61 | ,000 | ,791 | 61 | ,000 |
| motorin | ,304 | 61 | ,000 | ,812 | 61 | ,000 |
| 1. Lilliefors Significance Correction | | | | | | |

“**SPSS** bu testleri **%95 güven aralığında** yapmaktadır. Yani bizim değerimiz **0,05 anlamlılık düzeyi**nden büyük olduğu için verilerin **normale** **yakın** **dağıldığını** söyleyebiliriz.”

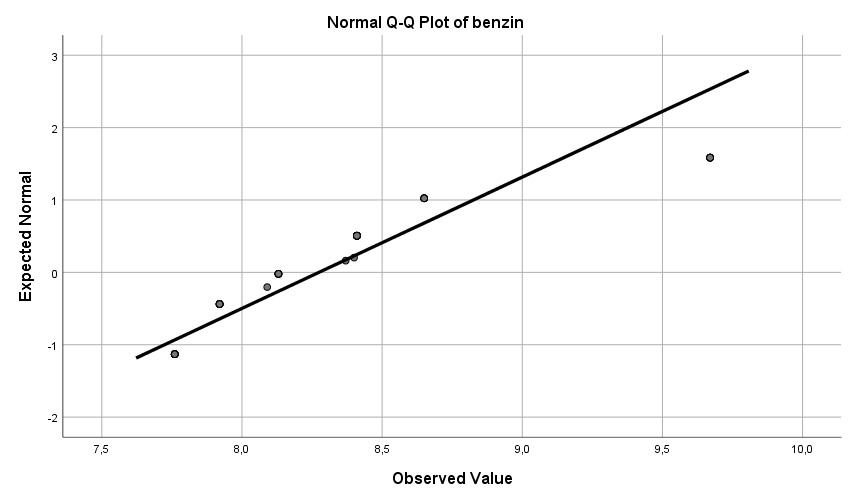
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | | | | |
|  | N | Mean | Std. Deviation | Minimum | Maximum | Percentiles | | |
| 25th | 50th (Median) | 75th |
| motorin | 61 | 8,1938 | ,65533 | 7,30 | 9,78 | 7,7450 | 8,1900 | 8,2200 |
| benzin | 61 | 8,2746 | ,55110 | 7,76 | 9,67 | 7,8400 | 8,1300 | 8,4100 |

|  |  |  |  |
| --- | --- | --- | --- |
| **One-Sample Kolmogorov-Smirnov Test** | | | |
|  | | motorin | benzin |
| N | | 61 | 61 |
| Normal Parametersa,b | Mean | 8,1938 | 8,2746 |
| Std. Deviation | ,65533 | ,55110 |
| Most Extreme Differences | Absolute | ,304 | ,206 |
| Positive | ,304 | ,206 |
| Negative | -,203 | -,175 |
| Test Statistic | | ,304 | ,206 |
| Asymp. Sig. (2-tailed) | | ,000c | ,000c |
| a. Test distribution is Normal. | | | |
| b. Calculated from data. | | | |
| c. Lilliefors Significance Correction. | | | |





Normallik için simetrik bi eğri olmasını bekleriz. Çan eğrisi şeklinde ama bakıldığında iki histogramı da çan eğrisine benzemiyor Q-Q grafiğine bakalım.



Noktalar çizgiye ne kadar yakınsa verilerin o kadar normale yakın oldupunu söylebiliriz



………………………………………………………………………………………………………………………………………………………

Ki Kare Testi

Normal dağılım parametreleri iki taneydi. Ortalamaya ve normal dağılıma baktım. Şimdi

Ki kare için serbest dağılım ve anlamlılık düzeyine bkacağım.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptive Statistics** | | | | | | | | |
|  | N | Mean | Std. Deviation | Minimum | Maximum | Percentiles | | |
| 25th | 50th (Median) | 75th |
| benzin | 61 | 8,2746 | ,55110 | 7,76 | 9,67 | 7,8400 | 8,1300 | 8,4100 |
| motorin | 61 | 8,1938 | ,65533 | 7,30 | 9,78 | 7,7450 | 8,1900 | 8,2200 |

|  |  |  |  |
| --- | --- | --- | --- |
| **benzin** | | | |
|  | Observed N | Expected N | Residual |
| 7,76 | 15 | 6,8 | 8,2 |
| 7,92 | 10 | 6,8 | 3,2 |
| 8,09 | 1 | 6,8 | -5,8 |
| 8,13 | 8 | 6,8 | 1,2 |
| 8,37 | 1 | 6,8 | -5,8 |
| 8,40 | 1 | 6,8 | -5,8 |
| 8,41 | 13 | 6,8 | 6,2 |
| 8,65 | 6 | 6,8 | -,8 |
| 9,67 | 6 | 6,8 | -,8 |
| Total | 61 |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **motorin** | | | |
|  | Observed N | Expected N | Residual |
| 7,30 | 7 | 8,7 | -1,7 |
| 7,53 | 8 | 8,7 | -,7 |
| 7,96 | 3 | 8,7 | -5,7 |
| 8,19 | 24 | 8,7 | 15,3 |
| 8,22 | 8 | 8,7 | -,7 |
| 8,72 | 5 | 8,7 | -3,7 |
| 9,78 | 6 | 8,7 | -2,7 |
| Total | 61 |  |  |

|  |  |  |
| --- | --- | --- |
| **Test Statistics** | | |
|  | benzin | motorin |
| Chi-Square | 32,393a | 33,443b |
| df | 8 | 6 |
| Asymp. Sig. | ,000 | ,000 |
| a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 6,8. | | |
| b. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 8,7. | | |

“Khi – Kare testi genel olarak aşağıdaki durumlarda kullanılmaktadır:

1. Sayısal olmayan değişkenler arasında bağımlılık olup olmadığının hipotez testleriyle araştırılmasında,

2. Örneklem sonucunda elde edilen frekans dağılımlarının herhangi bir olasılık dağılımına (Normal, Poisson, Üstel vb.) uyup uymadığının araştırılmasında,

3. İki ya da daha fazla örneğin aynı anakitleden seçilip seçilmediklerinin araştırılması (bağdaşıklık testi),

4. Anakitle varyanslarının tahminleri ve testlerinde.”

Benzin ve motorinin iki ay içerisindeki artışlarının karşılaştırılması gelen zamın iki veri grubu için değerlendirmesi amacıyla kullanmak istiyorum

Ki kare testi hipotezi

H0= Benzin ve motorinin iki ay içerisindeki artışı birbirinden bağımsızdır.

H1 = Benzin ve motorinin iki ay içerisindeki artışının birbirine bağımsız olduğu reddedilir.

Ki kare testi için düzgün bir hipotez oluşturamadım verimi ayarlayamadım. Sonuç olarak eksik kaldı.

…………………………………………………………………………………………….

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistics** | | | | |
|  | | benzin | motorin | ay |
| N | Valid | 61 | 61 | 61 |
| Missing | 0 | 0 | 0 |
| Variance | | ,304 | ,429 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **benzin** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 7,76 | 15 | 24,6 | 24,6 | 24,6 |
| 7,92 | 10 | 16,4 | 16,4 | 41,0 |
| 8,09 | 1 | 1,6 | 1,6 | 42,6 |
| 8,13 | 8 | 13,1 | 13,1 | 55,7 |
| 8,37 | 1 | 1,6 | 1,6 | 57,4 |
| 8,40 | 1 | 1,6 | 1,6 | 59,0 |
| 8,41 | 13 | 21,3 | 21,3 | 80,3 |
| 8,65 | 6 | 9,8 | 9,8 | 90,2 |
| 9,67 | 6 | 9,8 | 9,8 | 100,0 |
| Total | 61 | 100,0 | 100,0 |  |
|  |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **motorin** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 7,30 | 7 | 11,5 | 11,5 | 11,5 |
| 7,53 | 8 | 13,1 | 13,1 | 24,6 |
| 7,96 | 3 | 4,9 | 4,9 | 29,5 |
| 8,19 | 24 | 39,3 | 39,3 | 68,9 |
| 8,22 | 8 | 13,1 | 13,1 | 82,0 |
| 8,72 | 5 | 8,2 | 8,2 | 90,2 |
| 9,78 | 6 | 9,8 | 9,8 | 100,0 |
| Total | 61 | 100,0 | 100,0 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ay** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | ekim | 31 | 50,8 | 50,8 | 50,8 |
| kasım | 30 | 49,2 | 49,2 | 100,0 |
| Total | 61 | 100,0 | 100,0 |  |

benzin Stem-and-Leaf Plot

Frequency Stem & Leaf

15,00 77 . 666666666666666

,00 78 .

10,00 79 . 2222222222

1,00 80 . 9

8,00 81 . 33333333

,00 82 .

1,00 83 . 7

14,00 84 . 01111111111111

,00 85 .

6,00 86 . 555555

6,00 Extremes (>=9,67)

Stem width: ,10

Each leaf: 1 case(s)

………………………………………………………………………………………………………………..

motorin Stem-and-Leaf Plot

Frequency Stem & Leaf

15,00 Extremes (=<7,53)

3,00 79 . 666

,00 80 .

,00 80 .

,00 81 .

24,00 81 . 999999999999999999999999

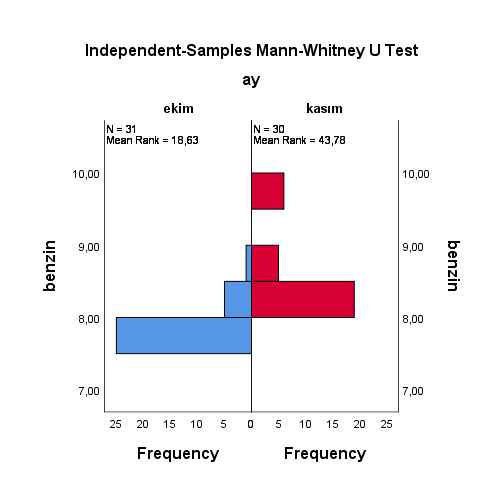
8,00 82 . 22222222

11,00 Extremes (>=8,72)

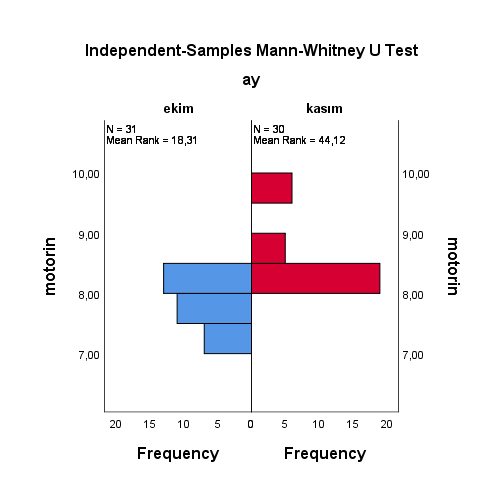
Stem width: ,10

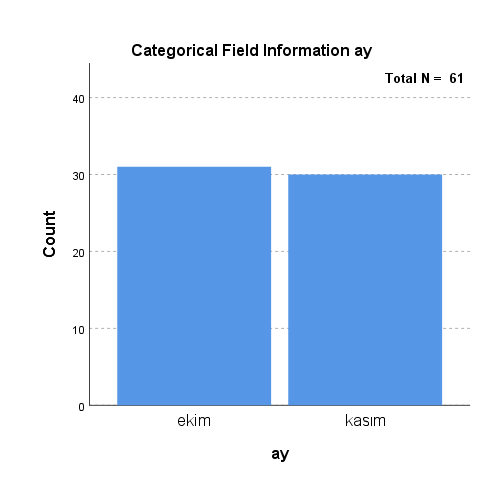
Each leaf: 1 case(s)

|  |  |
| --- | --- |
| **Independent-Samples Mann-Whitney U Test Summary** | |
| Total N | 61 |
| Mann-Whitney U | 848,500 |
| Wilcoxon W | 1313,500 |
| Test Statistic | 848,500 |
| Standard Error | 68,169 |
| Standardized Test Statistic | 5,626 |
| Asymptotic Sig.(2-sided test) | ,000 |



|  |  |
| --- | --- |
| **Independent-Samples Mann-Whitney U Test Summary** | |
| Total N | 61 |
| Mann-Whitney U | 858,500 |
| Wilcoxon W | 1323,500 |
| Test Statistic | 858,500 |
| Standard Error | 66,909 |
| Standardized Test Statistic | 5,881 |
| Asymptotic Sig.(2-sided test) | ,000 |





Aylık olarak genel durum. Ekim ayı kasım ayına göre fşyatları uygun olmasına rağmen demekki daha çok zam almış total daha yüksek çıkıyor.

……………………………………………………………………………………………

Bu testlerden sonra ay sutünuna ekim=1 kasım=2 sayılarını verdim string yapısından kurtarmak için. Mann whitney testini uyguladım.

**Mann-Whitney Test**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ranks** | | | | |
|  | ay | N | Mean Rank | Sum of Ranks |
| motorin | 1 | 31 | 18,31 | 567,50 |
| 2 | 30 | 44,12 | 1323,50 |
| Total | 61 |  |  |
| benzin | 1 | 31 | 18,63 | 577,50 |
| 2 | 30 | 43,78 | 1313,50 |
| Total | 61 |  |  |

|  |  |  |
| --- | --- | --- |
| **Test Statisticsa** | | |
|  | motorin | benzin |
| Mann-Whitney U | 71,500 | 81,500 |
| Wilcoxon W | 567,500 | 577,500 |
| Z | -5,881 | -5,626 |
| Asymp. Sig. (2-tailed) | ,000 | ,000 |
| a. Grouping Variable: ay | | |
|  | | |
|  | | |

1900002032 ESLEM NUR DOĞAN

KAYNAK

1)Benzin ve motorin değerleri için:

<https://www.opet.com.tr/akaryakit-fiyatlari-arsivi>

2)Normallik varsayımında değerleri 2,-2 arasında kabul edebilmek için kullandığım kaynak :

George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson

“A kurtosis value between ±1.0 is considered excellent for most psychometric purposes, but a value between ±2.0 is in many cases also acceptable, depending on the particular application." (George & Mallery (2012))

3) <https://dijilopedi.com> spss genel kullanım

4) Ozan Kocadağlı Kültür üniversitesi veri analizi ders notları