

FBE503 UYGULAMALI VERİ ANALİZİ DERSİ FİNAL RAPORU

Melike BEKTAŞ - 20424011001

21.06.2021

Özet

İstatistiksel analiz yöntemleri çoğu verinin analiz edilmesi için sıklıkla kullanılmaktadır. Bu test yöntemleri sonucunda elde edilen değerlerin yorumlanması önemli veri analizleri ve veri ilişkileri bulmamızı sağlamaktadır. Bu çalışmada farklı senaryolar üzerinden Sign Test, Korelasyon testi, Simple ve Multiple Lineer Regresyon, Wilcoxon Signed-Rank Testi, Mann Whitney Wilcoxon Testi ve Kruskall Wallis testi yapılmıştır. Elde edilen sonuç ve gözlemlere sonuçlar ve tartışma bölümünde yer verilmiştir.

1. GİRİŞ

Ülkemizde ve dünyada üretilen her bir ürünün satışa çıkmadan önce stoklarda belirli bir miktarda tutulması o ürünün bulunması ve kara borsaya düşmemesi için önem arz etmektedir. Şirketler stokta tutulan ürünlerin depolama maliyetini her geçen gün optimize etmek için yeni yöntemler bulmaya çalışmaktadır. Stokta bulunan ürünlerin fiyatlarını ve depolama maliyetlerini birden fazla faktör etkileyebilmektedir. Bu faktörlerden bazıları: ülkedeki işsizlik oranı, faiz oranları, ülkedeki enflasyon ve ekonomi durumudur. Bu çalışmada stok fiyatlarına faiz oranlarının etkisi simple lineer regresyon modeli oluşturularak incelenmiştir. İnsanların yaşları ve kan değerlerinde bulunan glikoz oranlarının ilişkisi de korelasyon testi uygulanarak test edilmiş ve değerlendirilmiştir. Bir bilgisayar firmasının yeni üretmiş olduğu bilgisayarın aynı firmanın en çok satan bilgisayar ürünü ile piyasada yarışıp yarışamayacağı Sign Testi ile test edilmiştir. Bir sarkacın ısı uygulandıktan önce ve sonraki ölçüm değerleri karşılaştırılmak istenmiştir. Kullanılan veriler normal dağılmadığı ve eşli (paired) olduğundan dolayı Wilcoxon Signed-Rank Testi uygulanmıştır. Bir okuldaki öğrencilerin İngilizce ve Matematik sınavlarından almış oldukları notlar karşılaştırılmak istenmiştir. Kullanılan veriler normal dağılmadığı için Mann Whitney Wilcoxon Testi uygulanmıştır. Birbirinden farklı A, B, C, D, E, F, G ve H ilaç firmalarının böcek ölümlerine olan etkisi Kruskall Wallis Test kullanılarak test edilmiştir. Çalışmada ayrıca winequality veri seti kullanılarak sabit asit miktarı ve uçucu asit miktarının şarap kalitesine olan etkisi incelenmiş ve bunun için multiple lineer regresyon modeli uygulanmıştır.

2. PROBLEMİN TANIMI

Faiz oranlarının bir şirketin stok fiyatları üzerine etkisini temsil edecek simple lineer regresyon modelini oluşturunuz. Simple lineer regresyon modeli için Stock_Index_Price bağımlı değişkeni ile Interest_Rate bağımsız değişkeni kullanılmıştır.

```
# A tibble: 6 x 5
  Year Month Interest_rate Unemployed_rate Stock_Index_Price
<dbl> <dbl> <dbl> <dbl> <dbl>
1 2017    12      2.75      5.3      1464
2 2017    11      2.5      5.3      1394
3 2017    10      2.5      5.3      1357
4 2017     9      2.5      5.3      1293
5 2017     8      2.5      5.4      1256
6 2017     7      2.5      5.6      1254
```

Şekil 1. Ay, Yıl, Faiz oranı ve İşsizlik Oranı ile Elde Edilen Stok Fiyatları

Kan değerlerinden glikoz seviyesinin yaş ile olan ilişkisini korelasyon testi kullanarak inceleyiniz.

```
# A tibble: 6 x 2
  Age `Glucose Level`
<dbl> <dbl>
1 43 99
2 21 65
3 25 79
4 42 75
5 57 87
6 59 81
```

Şekil 2. Yaş ve Glikoz Seviyesi Değerleri

Bir sarkacın belirli bir sıcaklık değeri uygulandıktan sonraki ölçümlerinin istatistiksel olarak değerlendiriniz. Dv0 ve Dv6 eşli verileri kullanılmıştır.

```
# A tibble: 6 x 2
  dv0 dv6
<dbl> <dbl>
1 296 175
2 376 329
3 309 238
4 222 60
5 100 171
6 316 291
```

Şekil 3. Bir Sarkacın Belirli Aralıklardaki Uzunluğunun Ölçüm Değerleri

Bir okuldaki öğrencilerin Matematik ve İngilizce derslerinden almış oldukları notların istatistiksel olarak karşılaştırınız.

	Matematik	İngilizce
1	2	10
2	100	50
3	85	90
4	15	45
5	45	60
6	96	1

Şekil 4. Öğrencilerin Matematik ve İngilizce Sınavlarından Almış Oldukları Notlar

A, B, C, D, E, F, G ve H isimli böcek ilacı markalarının böceklerin azalmasına olan etkisini karşılaştırınız.

```
> head(OrchardSprays)
```

	x	decrease	rowpos	colpos	treatment
1	1	57	1	1	D
2	2	95	2	1	E
3	3	8	3	1	B
4	4	69	4	1	H
5	5	92	5	1	G
6	6	90	6	1	F

Şekil 5. A, B, C, D, E, F, G ve H isimli böcek ilacı markalarının karşılaştırılması için kullanılan veri seti

Şarap kalitesine sabit asit ve uçucu asit miktarının etkisini inceleyiniz. Multiple lineer regresyon modeli oluşturunuz. Multiple lineer regresyon modeli için quality bağımlı değişkeni, fixed.acidity ile volatile.acidity bağımsız değişkenleri kullanılmıştır.

```
> head(winequality.red)
```

	fixed.acidity	volatile.acidity	citric.acid	residual.sugar	chlorides
1	7.4	0.70	0.00	1.9	0.076
2	7.8	0.88	0.00	2.6	0.098
3	7.8	0.76	0.04	2.3	0.092
4	11.2	0.28	0.56	1.9	0.075
5	7.4	0.70	0.00	1.9	0.076
6	7.4	0.66	0.00	1.8	0.075

	free.sulfur.dioxide	total.sulfur.dioxide	density	pH	sulphates	alcohol	quality
1	11	34	0.9978	3.51	0.56	9.4	5
2	25	67	0.9968	3.20	0.68	9.8	5
3	15	54	0.9970	3.26	0.65	9.8	5
4	17	60	0.9980	3.16	0.58	9.8	6
5	11	34	0.9978	3.51	0.56	9.4	5
6	13	40	0.9978	3.51	0.56	9.4	5

Şekil 6. Wine Quality Veri Seti

3. METOT

Bu çalışmada Sign Test, Simple Linear Regresyon, Multiple Linear Regresyon, Wilcoxon Signed-Rank Test, Mann Whitney Wilcoxon Test ve Kruskall Wallis Testi uygulanmıştır. Araştırma kapsamında stok fiyatları için simple lineer regresyon yöntemi ile stok fiyatlarını temsil eden bir model oluşturulmuştur. Glikoz seviyesi ve yaş arasındaki ilişkinin tespit edilebilmesi için korelasyon testi yapılmıştır. Yeni çıkan bir bilgisayarın piyasada tutup tutmayacağını test etmek için Sign Testi uygulanmıştır. Bir sarkacın sıcaklık uygulandıktan önceki ve sonraki ölçümlerinin karşılaştırılması için Wilcoxon Signed-Rank Test uygulanmıştır. Öğrencilerin Matematik ve İngilizce sınavlarından almış oldukları notları istatistiksel olarak karşılaştırmak için Mann Whitney Wilcoxon Test, farklı ilaç firmalarının böceklerin azalmasına olan etkisini incelemek için ise Kruskall Wallis Testi uygulanmıştır. Sabit asit ve uçucu asit miktarının şarap kalitesine etkisinin test edilebilmesi için multiple lineer regresyon modeli oluşturulmuştur. Simple lineer regresyon ve multiple lineer regresyon için iki adet hipotez kurulur. Bunlar:

H0: Geliştirilen modeldeki beta katsayısı sıfırdır.

H1: Geliştirilen modeldeki beta katsayısı sıfırdan farklı bir sayıdır.

Korelasyon testi için iki adet hipotez kurulur. Bunlar:

H0: $r = 0$. Korelasyon yoktur.

H1: $r \neq 0$. Korelasyon ilişkisi bulunmaktadır.

Sign testi için iki adet hipotez kurulur. Bunlar:

H0: Medyan değeri iddaa edilen değere eşittir.

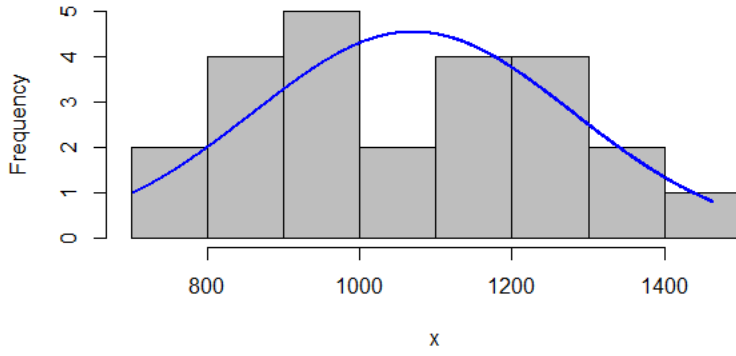
H1: Medyan değeri iddaa edilen değere eşit değildir.

Wilcoxon Test için iki adet hipotez kurulur. Bunlar:

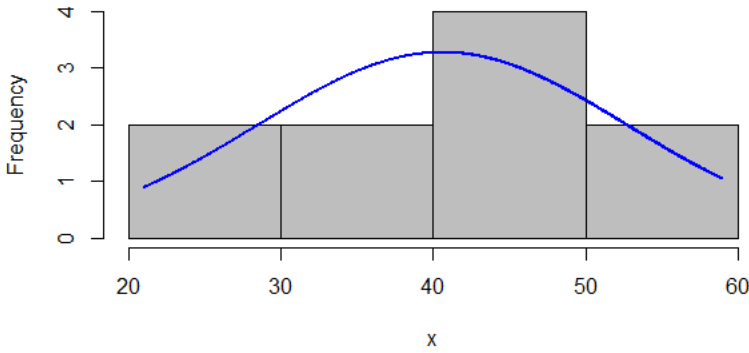
H0: İki örneğin de medyan değerleri eşittir.

H1: İki örneğin medyan değerleri birbirinden farklıdır.

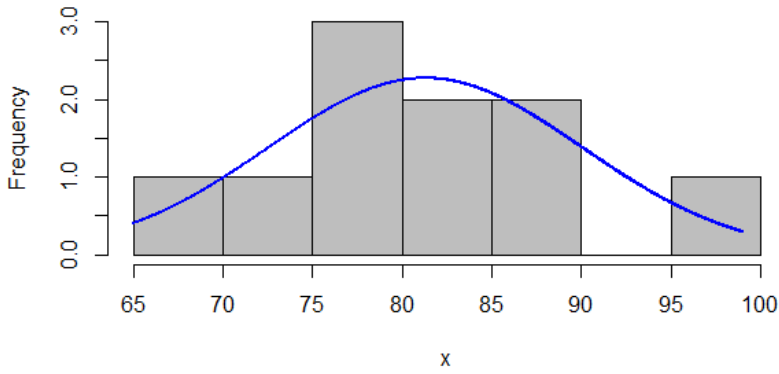
Verilerin excel dosyasından alınabilmesi için “readxl” kütüphanesinin yüklenerek verilerin erişilebilir olması için “attach” komutu kullanılmalıdır. Veriler attach yapıldıktan sonra simple lineer regresyon analizi için bağımlı değişkenimiz olan “Stock_Index_Price” değişkenine ve korelasyon analizi için kullanılan “Age” ve “Glucose Level” değişkenlerine normal dağılım (shapiro) testi yapılmıştır. Ayrıca Wilcoxon Signed-Rank Testi, Mann Whitney Wilcoxon Testi ve Kruskall Wallis testleri için kullanılan veri setlerindeki değişkenlere de shapiro testi yapılmış ve verilen normal dağılmadığı ve transforme de edilemedikleri görülmüştür.



Şekil 7. “Stock_Index_Price” Bağımlı Değişkeninin Histogramı ve İdeal Dağılım Eğrisi



Şekil 8. “Age” Değişkeni Histogramı ve İdeal Dağılım Eğrisi



Şekil 9. “Glucose Level” Değişkeni Histogramı ve İdeal Dağılım Eğrisi

Simple lineer regresyon analizi için kullandığımız “Stock_Index_Price” bağımlı değişkenimizin shapiro testi sonucu p-value değeri: 0.7201

Korelasyon testi için kullandığımız “Age” değişkeni shapiro testi sonucu p-value değeri: 0.7117

Korelasyon testi için kullandığımız “Glucose Level” değişkeni shapiro testi sonucu p-value değeri: 0.6872

Shapiro test sonuçlarında hem simple lineer regresyon testi için kullandığımız “Stock_Index_Price” değişkeninin hem de korelasyon analizi testi için kullandığımız “Age” ve “Glucose Level” değişkenlerimizin normal dağıldıkları ve ilgili analizler için uygun oldukları görülmüştür. Modeller oluşturulmuştur.

Simple Linear Regresyon Analizi

```
> shapiro.test(Stock_Index_Price)

      shapiro-wilk normality test

data:  Stock_Index_Price
W = 0.97215, p-value = 0.7201

> model<-lm(Stock_Index_Price~Interest_rate)
> plot(Stock_Index_Price~Interest_rate)
> abline(model)
> summary(model)

Call:
lm(formula = Stock_Index_Price ~ Interest_rate)

Residuals:
    Min       1Q   Median       3Q      Max
-183.892  -30.181    4.455   56.608  101.057

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -99.46     95.21   -1.045    0.308
Interest_rate   564.20     45.32  12.450 1.95e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 75.96 on 22 degrees of freedom
Multiple R-squared:  0.8757,    Adjusted R-squared:  0.8701
F-statistic: 155 on 1 and 22 DF,  p-value: 1.954e-11
```

Simple Linear Regresyon Power Analizi

```
> pwr.f2.test (u =1, v =22, f2 =6.69, sig.level =0.05, power = NULL)

      Multiple regression power calculation

              u = 1
              v = 22
             f2 = 6.69
    sig.level = 0.05
         power = 1
```

Korelasyon Testi Analizi

```
> shapiro.test(Age)

      shapiro-wilk normality test

data:  Age
W = 0.95364, p-value = 0.7117

> plotNormalHistogram(Age)
> shapiro.test(`Glucose Level`)

      shapiro-wilk normality test

data:  Glucose Level
W = 0.95158, p-value = 0.6872

> cor.test(Age, `Glucose Level`, method = "pearson")

      Pearson's product-moment correlation

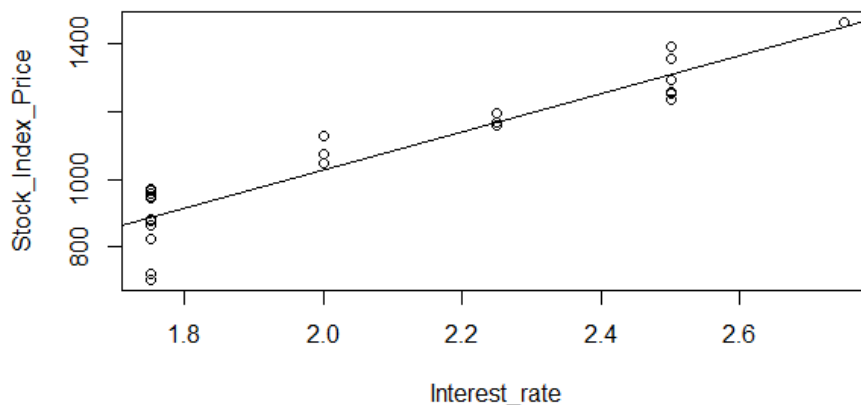
data:  Age and Glucose Level
t = 1.7945, df = 8, p-value = 0.1105
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1416869  0.8714191
sample estimates:
      cor 
0.5357319
```

Korelasyon Testi Power Analizi

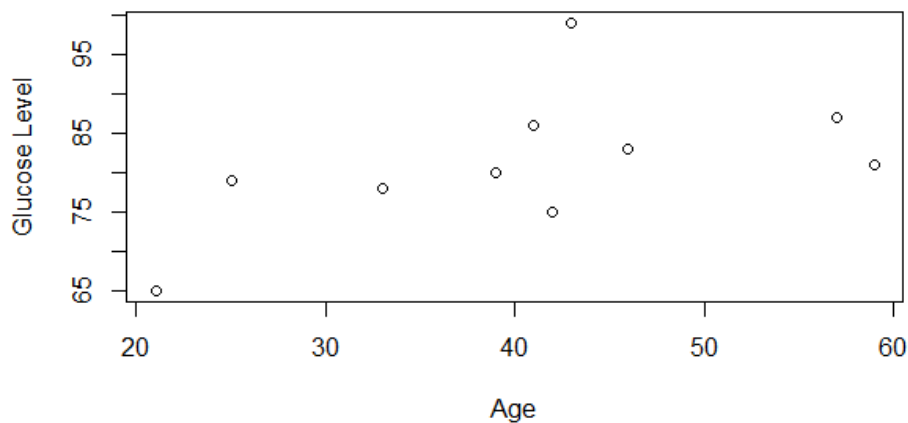
```
> library(pwr)
> pwr.r.test(n = 10, r =0.5357319 , sig.level =0.05 , power =NULL )

      approximate correlation power calculation (arctangh transformation)

      n = 10
      r = 0.5357319
sig.level = 0.05
power = 0.3789391
alternative = two.sided
```



Şekil 10. Simple Linear Regresyon Sonucu Elde Edilen Grafik



Şekil 11. Korelasyon Testi Sonucu Elde Edilen Grafik

Multiple Lineer Regresyon Analizi

```
> attach(winequality.red)
> mymodel<-lm(quality~fixed.acidity+volatile.acidity)
> summary(mymodel)
```

Call:
lm(formula = quality ~ fixed.acidity + volatile.acidity)

Residuals:

Min	1Q	Median	3Q	Max
-2.81291	-0.54009	-0.00973	0.46502	2.95433

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.45088	0.12121	53.221	<2e-16 ***
fixed.acidity	0.01192	0.01105	1.079	0.281
volatile.acidity	-1.73175	0.10747	-16.113	<2e-16 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7436 on 1596 degrees of freedom
Multiple R-squared: 0.1532, Adjusted R-squared: 0.1521
F-statistic: 144.3 on 2 and 1596 DF, p-value: < 2.2e-16

```
> library(MASS)
> stepAIC(mymodel, direction = "backward")
Start: AIC=-944.31
quality ~ fixed.acidity + volatile.acidity
```

	Df	Sum of Sq	RSS	AIC
- fixed.acidity	1	0.643	883.20	-945.14
<none>			882.55	-944.31
- volatile.acidity	1	143.573	1026.13	-705.29

Step: AIC=-945.14
quality ~ volatile.acidity

	Df	Sum of Sq	RSS	AIC
<none>			883.2	-945.14
- volatile.acidity	1	158.97	1042.2	-682.50

Call:
lm(formula = quality ~ volatile.acidity)

Coefficients:

(Intercept)	6.566
volatile.acidity	-1.761


```

> bptest(mymodel)

studentized Breusch-Pagan test

data: mymodel
BP = 1.1777, df = 2, p-value = 0.555

> cor.test(fixed.acidity,volatile.acidity)

Pearson's product-moment correlation

data: fixed.acidity and volatile.acidity
t = -10.589, df = 1597, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3013681 -0.2097433
sample estimates:
      cor
-0.2561309

wine_quality = 0.01*fixed.acidity-1.73*volatile.acidity+6.45

mymodel<-lm(quality~fixed.acidity+volatile.acidity+fixed.acidity:volatile.acidity
            +volatile.acidity:fixed.acidity)
stepAIC(mymodel, direction = "backward")

```

Sign Testi

Toplamda 25 kişiye yeni bilgisayar ile ilgili fikirleri sorulmuş ve 25 kişiden 10 kişiden olumlu 15 kişiden ise olumsuz geri dönüş alınmıştır.

```

> binom.test(10,25)

Exact binomial test

data: 10 and 25
number of successes = 10, number of trials = 25, p-value = 0.4244
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.2112548 0.6133465
sample estimates:
probability of success
              0.4

```

Wilcoxon Signed-Rank Testi

```

> mydat<-attach(wilcoxon_test_data)
> shapiro.test(dv0)

shapiro-wilk normality test

data: dv0
W = 0.86249, p-value = 0.003762

> shapiro.test(dv6)

shapiro-wilk normality test

data: dv6
W = 0.86497, p-value = 0.0042

```

```
> transformTukey(dv0)

      lambda      w Shapiro.p.value
423    0.55 0.8924          0.01488

if (lambda > 0){TRANS = x ^ lambda}
if (lambda == 0){TRANS = log(x)}
if (lambda < 0){TRANS = -1 * x ^ lambda}

[1] 22.86707 26.08275 23.41408 19.52064 12.58925 23.70434 23.90990 28.68594
[9] 19.42372 26.04457 23.45573 36.99918 22.26573 23.74557 25.54405 24.59825
[17] 21.42771 12.58925 24.23582 12.02495 21.96003 15.73440 23.82785 23.03651
> transformTukey(dv6)

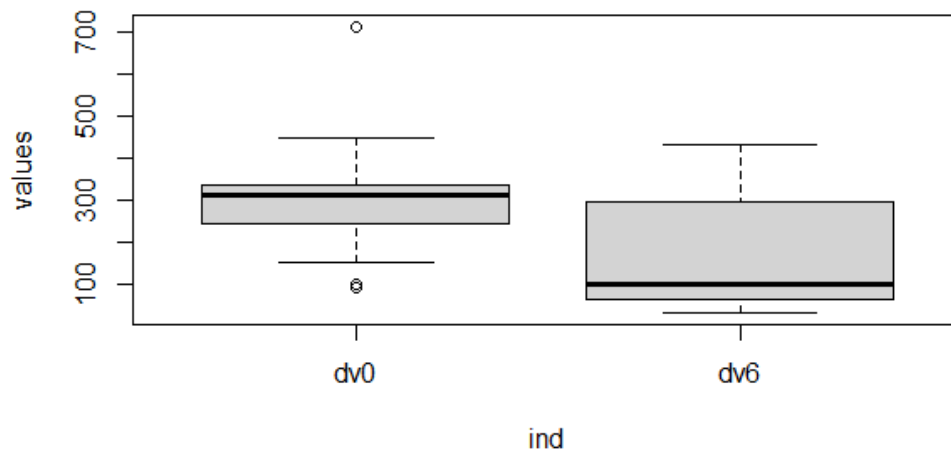
      lambda      w Shapiro.p.value
395   -0.15 0.9137          0.04238

if (lambda > 0){TRANS = x ^ lambda}
if (lambda == 0){TRANS = log(x)}
if (lambda < 0){TRANS = -1 * x ^ lambda}

[1] -0.4608338 -0.4191994 -0.4400616 -0.5410997 -0.4624349 -0.4269884 -0.4128904
[8] -0.4024180 -0.5287316 -0.4180645 -0.4250420 -0.4381523 -0.5126550 -0.5974420
[15] -0.4982474 -0.4186290 -0.5058606 -0.5794720 -0.5384449 -0.5772186 -0.5058606
[22] -0.5595177 -0.5310356 -0.5794720
> wilcox.test(dv0, dv6, paired = TRUE)

      wilcoxon signed rank test with continuity correction

data:  dv0 and dv6
V = 283.5, p-value = 0.0001446
alternative hypothesis: true location shift is not equal to 0
```



Şekil 12. Wilcoxon Signed-Rank Testi Sonucu Elde Edilen Grafik

Mann Whitney Wilcoxon Testi

```
> notlar<-attach(notlar)
> shapiro.test(Matematik)
```

shapiro-wilk normality test

```
data: Matematik
W = 0.84886, p-value = 0.02755
```

```
> shapiro.test(ingilizce)
```

shapiro-wilk normality test

```
data: ingilizce
W = 0.85772, p-value = 0.03595
```

```
> transformTukey(Matematik)
```

```
lambda      w shapiro.p.value
493      2.3 0.8568          0.03492
```

```
if (lambda > 0){TRANS = x ^ lambda}
if (lambda == 0){TRANS = log(x)}
if (lambda < 0){TRANS = -1 * x ^ lambda}
```

```
[1] 4.924578 39810.717055 27394.509523 507.002261 6344.374073 36242.974709
[7] 28899.762986 34529.800032 40.516415 39810.717055 1796.547484 25221.555093
[13] 13755.717993
```

```
> transformTukey(ingilizce)
```

```
lambda      w shapiro.p.value
435      0.85 0.8585          0.03677
```

```
if (lambda > 0){TRANS = x ^ lambda}
if (lambda == 0){TRANS = log(x)}
if (lambda < 0){TRANS = -1 * x ^ lambda}
```

```
[1] 7.079458 27.805103 45.825386 25.423225 32.465983 1.000000 3.927575 45.392229
[9] 49.265417 49.265417 50.118723 12.760729 45.392229
```

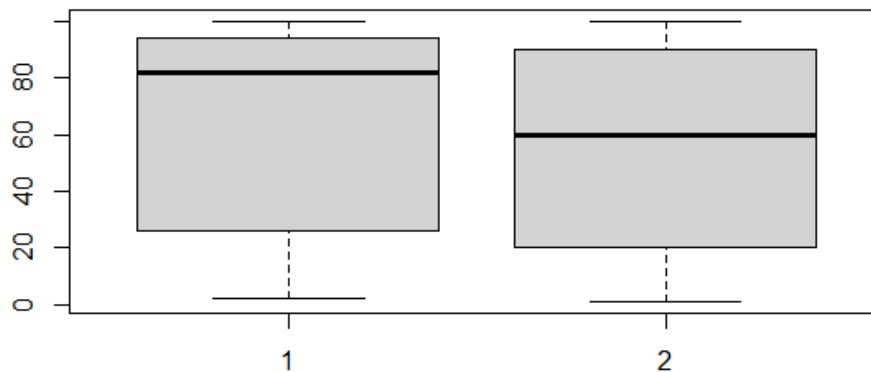
```
> wilcox.test(Matematik,ingilizce)
```

wilcoxon rank sum test with continuity correction

```
data: Matematik and ingilizce
```

```
W = 87, p-value = 0.9182
```

```
alternative hypothesis: true location shift is not equal to 0
```



Şekil 13. Mann Whitney Wilcoxon Testi Sonucu Elde Edilen Grafik

Kruskall Wallis Testi

```
> attach(OrchardSprays)
> shapiro.test(decrease)
```

shapiro-wilk normality test

```
data: decrease
W = 0.91892, p-value = 0.0004483
```

```
> transformTukey(decrease)
```

```
lambda      W shapiro.p.value
420  0.475 0.9427          0.005057
```

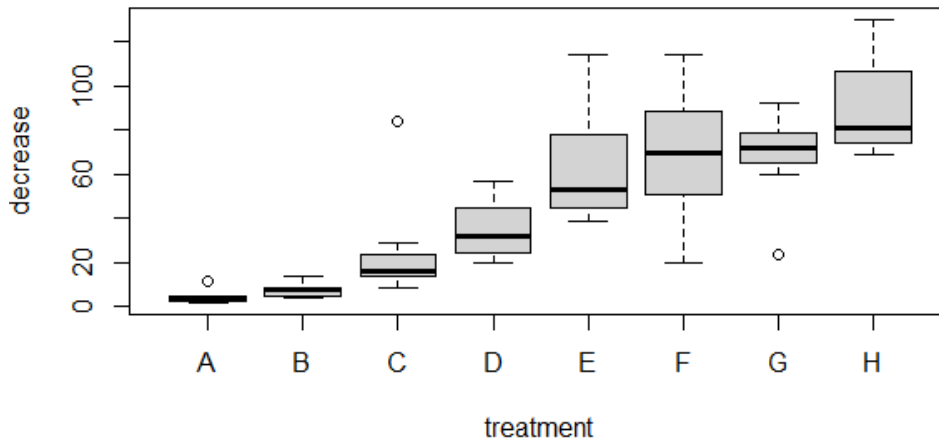
```
if (lambda > 0){TRANS = x ^ lambda}
if (lambda == 0){TRANS = log(x)}
if (lambda < 0){TRANS = -1 * x ^ lambda}
```

```
[1] 6.824024 8.697986 2.685145 7.472283 8.566417 8.477449 3.619456 1.389918
[9] 8.204132 2.342188 9.984050 5.485847 6.472854 1.389918 7.472283 7.574391
[17] 8.342028 7.624879 2.147884 5.698436 4.341610 3.732132 7.624879 1.931873
[25] 10.095390 1.931873 9.484836 2.839652 4.149438 4.524810 2.985383 6.472854
[33] 5.968944 4.868555 6.992329 2.147884 3.841167 2.520122 8.063626 7.574391
[41] 3.255450 4.950386 6.034482 7.871964 1.931873 4.785175 6.226536 7.823236
[49] 2.685145 7.624879 3.381607 6.824024 1.931873 8.063626 4.149438 7.047445
[57] 8.016185 9.484836 5.698436 3.502764 8.296344 6.709224 1.685127 4.049561
```

```
> kruskal.test(decrease~treatment, data = OrchardSprays)
```

kruskal-wallis rank sum test

```
data: decrease by treatment
kruskal-wallis chi-squared = 48.874, df = 7, p-value = 2.402e-08
```



Şekil 14. Kruskall Wallis Testi Sonucu Elde Edilen Grafik

```

> kruskalmc(decrease-treatment, Orchardsprays, probs=0.05)
Multiple comparison test after Kruskal-wallis
p.value: 0.05
Comparisons
  obs.dif critical.dif difference
A-B  5.0000    29.08039    FALSE
A-C 18.3750    29.08039    FALSE
A-D 23.8750    29.08039    FALSE
A-E 35.9375    29.08039     TRUE
A-F 39.0625    29.08039     TRUE
A-G 40.0625    29.08039     TRUE
A-H 47.6875    29.08039     TRUE
B-C 13.3750    29.08039    FALSE
B-D 18.8750    29.08039    FALSE
B-E 30.9375    29.08039     TRUE
B-F 34.0625    29.08039     TRUE
B-G 35.0625    29.08039     TRUE
B-H 42.6875    29.08039     TRUE
C-D  5.5000    29.08039    FALSE
C-E 17.5625    29.08039    FALSE
C-F 20.6875    29.08039    FALSE
C-G 21.6875    29.08039    FALSE
C-H 29.3125    29.08039     TRUE
D-E 12.0625    29.08039    FALSE
D-F 15.1875    29.08039    FALSE
D-G 16.1875    29.08039    FALSE
D-H 23.8125    29.08039    FALSE
E-F  3.1250    29.08039    FALSE
E-G  4.1250    29.08039    FALSE
E-H 11.7500    29.08039    FALSE
F-G  1.0000    29.08039    FALSE
F-H  8.6250    29.08039    FALSE
G-H  7.6250    29.08039    FALSE

```

4. SONUÇLAR VE TARTIŞMA

Sonuç olarak yaş ve glikoz seviyesi bağımlı değişkenlerinin Pearson korelasyon testi sonucunda p-value değeri 0.1105 olarak bulunmuştur. P-value değeri 0.05 değerinden büyük olduğu için korelasyon testi null hipotezi kabul edilmiş ve alternatif hipotez reddedilmiştir. Simple lineer regresyon analizi sonucunda Adjusted R-squared değeri 0.8701 olarak bulunmuştur. Bu sonuç stok fiyatları bağımlı değişkeni ile faiz oranları bağımsız değişkeninin birbiri ile ilişkili olduğunu göstermiştir. Sign testi sonucunda P-value değeri 0.4244 olarak bulunmuştur. Bu sonuç bize yeni üretilen bilgisayarında piyasada en çok satış yapan aynı markalı bilgisayar ile yarışabileceğini, yeni bilgisayarında kabul gördüğünü göstermiştir. Bir sarkacın ısı verildikten önce ve sonraki uzunluk değişimleri Wilcoxon Signed-Rank Testi ile karşılaştırılmıştır. Test sonucunda P-value değeri 0.0001446 bulunmuştur. Bu değer sonucunda ısı verildikten sonraki sarkaç uzunluğu ile önceki sarkaç uzunluğunun istatistiksel olarak birbirinden farklı olduğu gözlemlenmiştir. Öğrencilerin Matematik ve İngilizce sınavlarından almış oldukları notları karşılaştırmak için Mann Whitney Wilcoxon Testi uygulanmış ve testin sonucunda 0.9182 P-value değeri elde edilmiştir. Bu sonuç iki sınavında istatistiksel olarak aynı olduğunu göstermiştir. Böceklerin azalması için kullanılan ilaçları incelemek için Kruskal Wallis testi uygulanmıştır ve elde edilen 2.402e-08 P-value değerinden dolayı istatistiksel olarak birbileri ile aynı olmadıkları sonucuna ulaşılmıştır. Multiple Lineer Regresyon analizi sonucunda Multiple R-squared değeri 0.1532 olarak bulunmuştur ve bu sonuç bize bağımsız değişkenlerimizin şarap kalitesi üzerindeki etkisinin az olduğunu göstermiştir.