

# Nonparametrik İstatistikler \*

RStudio İle Temel İstatistiksel Uygulamalar ve Raporlama Eğitimi

*Dr. Öğrt. Üyesi Eren Halil ÖZBERK*

*Trakya Üniversitesi Eğitimde Ölçme ve Değerlendirme ABD*

*erenozberk@gmail.com*

## 1. Nonparametrik İstatistikler

Hipotez testleri oluşturulurken, herhangi bir evrenden çekilen örneklem dağılımı yardımıyla evren dağılımı hakkında birtakım bilgilere sahip olunabilir. Evren dağılımı bilindiği durumlarda parametrik istatistikler (t-testi, ANOVA, ANCOVA...) kullanılır. Ancak evren dağılımı bilinmediği ya da herhangi bir varsayımda bulunulmadığı durumlarda nonparametrik testler kullanılır.

### 1.1. Varsayımlar

Nonparametrik testler, parametrik testlerin varsayımlarının karşılanmaması durumunda kullanılan istatistiksel yöntemlerdir. Parametrik ve nonparametrik testlerin varsayımları aşağıdaki gibidir:

- Parametrik testlerde örneklem veya örneklemelerin çekildikleri evrenlerin **normal dağılım göstermesi** ve **varyanslarının homojen olması** istenir. Nonparametrik testlerde ise evren dağılımı konusunda bir varsayım bulunmamaktadır.
- Parametrik testlerde örneklemdeki **gözlem sayılarının 30 veya 30 üzeri olması** istenirken, nonparametrik testler **gözlem sayısının 30'dan az** olduğu durumlarda kullanılabilir. Gözlem sayısının 30 olarak belirlenmesi örneklem istatistiklerinin genellikle 30'a yakinken normal dağılım gösterdiği varsayımdır. Ancak 30 ve üzerinde olup normal dağılım göstermeyen durumlarla da karşılaşılabilir. Bu bakımdan sadece örneklem sayısına bakarak nonparametrik test yapmaya karar vermek hatalı olabilir.
- Parametrik testlerde verilerin en az **eşit aralık** düzeyinde ve **sürekli** olması varsayımı vardır. Nonparametrik testlerde ise veri tipi en az **sıralama** düzeyindedir ve süreklilik belirtmeyebilir.
- Hem parametrik hem de nonparametrik testlerde örneklem evrenden **rastgele** seçilmelidir.

Nonparametrik testlerde verilerin sıralama düzeyinde kullanılmasını şöyle açıklayalım. Örneğin A sınıfındaki öğrencilerin boy uzunluklarını içeren bir veri üretelim.

```
set.seed(15)
boy = sample(160:180, 14, replace = TRUE)
names(boy) = paste("ID", 1:14, sep = "")
boy
```

```
## ID1 ID2 ID3 ID4 ID5 ID6 ID7 ID8 ID9 ID10 ID11 ID12 ID13 ID14
## 172 164 180 173 167 180 177 165 174 177 162 173 170 174
```

```
sort(boy)
```

```
## ID11 ID2 ID8 ID5 ID13 ID1 ID4 ID12 ID9 ID14 ID7 ID10 ID3 ID6
## 162 164 165 167 170 172 173 173 174 174 177 177 180 180
```

```
rank(boy)
```

```
## ID1 ID2 ID3 ID4 ID5 ID6 ID7 ID8 ID9 ID10 ID11 ID12 ID13 ID14
## 6.0 2.0 13.5 7.5 4.0 13.5 11.5 3.0 9.5 11.5 1.0 7.5 5.0 9.5
```

\*Bu belgede kullanılan örnekler Şener Büyükoztürk'ün *Sosyal Bilimler için Veri Analizi El Kitabı* isimli kaynağından yazarın izni ile alınmıştır.

```
sort(rank(boy))
```

```
## ID11 ID2 ID8 ID5 ID13 ID1 ID4 ID12 ID9 ID14 ID7 ID10 ID3 ID6
## 1.0 2.0 3.0 4.0 5.0 6.0 7.5 7.5 9.5 9.5 11.5 11.5 13.5 13.5
```

Dikkat edilirse 11 numaralı öğrencinin en kısa boya sahip olduğu (162 cm), 6 numaraalı öğrencinin ise en uzun boya sahip olduğu (180 cm) görülmüştür.

Nonparametrik testler gruplar arasındaki sistematik farklılıkları karşılaştırmak için kullanılır. Parametrik testlerde genelde ortalamalar karşılaştırılırken, nonparametrik testlerde Karşılaştırmalar genellikle ortanca değere, verinin şekline ve dağılımına göre yapılır. Bu bakımdan nonparametrik testler ile ilgili araştırmalarda dikkat edilmesi gereken bazı noktalar vardır. - Nonparametrik testler oluşturulurken hipotezlerin ve yorumlamaların ortalamalar üzerine kurulmamasına dikkat etmek gerekir. Elde edilen fark dağılımlara ait farktır ve ortalamaları yorumlamak mümkün değildir. - Veriler sıralama ölçeği düzeyinde olduğu için yorumlanması eşit aralıklı ölçekteki gibi yapılmamalıdır. - Nonparametrik testler, parametrik testlere göre genellikle güçlü değildir. Bu bakımdan etki büyüklüklerinin ve güç analizlerinin verilmesi önemlidir.

## 2. Mann Whitney-U Testi

İki ilişkisiz örneklemden elde edilen puanların karşılaştırılmasında kullanılır. Parametrik testlerdeki **bağımsız örneklem t-testi** ile benzerdir. Mann Whitney-U testinde iki grup birbirinden bağımsızdır ve bağımlı değişkenler sıralama düzeyindedir. İki Örneklem Wilcoxon İşaretli Sıralar Testi olarak da adlandırılır.

Mann Whitney-U testinde eşit aralıklı düzeyde veri olmadığından **ortalamalar** yerine **dağılımlar** karşılaştırılır. Mann Whitney-U testinde hipotezler ilişkisiz grupların ortancaları üzerine değil bir gruptaki gözlemlerin diğer gruptaki gözlemlerden büyük/küçük/farklı olduğu üzerine kurulur.

Mann Whitney-U analizleri `wilcox.test()` fonksiyonu ile yürütülebilir. Fonksiyonunun kullanımı aşağıdaki gibidir:

```
wilcox.test(x, y = NULL,
            alternative = c("two.sided", "less", "greater"),
            conf.int = FALSE, conf.level = 0.95, ...)
```

`wilcox.test()` fonksiyonunun argümanları Tablo 2.1’de açıklanmıştır.

**Tablo 2.1** `wilcox.test()` fonksiyonu argümanları ve açıklamaları

Argümanlar Açıklamalar	
x	Birinci örnekleme ait sayısal vektör
y	İkinci örnekleme ait sayısal vektör
alternative	Hipotezin özelliğini belirtir. İki yönlü için “two.sided”, büyüktür “greater”, küçüktür “less”
conf.int	Güven aralığı kestirimleri istenirse “TRUE”
conf.level	Güven aralığının düzeyini belirler. Halihazırda 0.95 olarak atanmıştır.
paired	Ölçümlerin ilişkisini belirler. Halihazırda “FALSE” olarak kayıtlıdır. “TRUE” seçilir.

### 2.1. Testin Aşamaları

#### Örnek

Araştırmacı, anti-sosyal davranış gösteren ilkökul öğrencilerine iletişim becerileri kazandırmaya yönelik bir programın etkililiğini incelemek istiyor. Bu tür davranışları gözlenen 14 öğrenci yansız olarak iki gruba (deney-kontrol) ayrılıyor. Deney grubu olarak atanan öğrenciler, iletişim becerilerini arttırmaya yönelik dört haftalık X programını alırken, diğer öğrenciler üzerinde hiçbir işlem yapılmıyor. Dört haftalık süre sonunda

öğrencilerin iletişim becerileri, geliştirilen bir araçla ölçülüyor. Yüksek puan, iletişim becerilerinin yüksek olduğunu göstermektedir. Çalışmaya ilişkin veriler “iletirim.csv” dosyasına kaydedilmiştir.

## 2.2. Araştırma Problemi ve Hipotezlerin Oluşturulması

**Araştırma Problemi** X programına katılan öğrencilerin iletişim becerileri, böyle bir programa katılmayan öğrencilere göre anlamlı bir şekilde daha yüksek midir?

**Sıfır Hipotezi**  $H_0$  : X programına katılan öğrencilerin iletişim becerileri, böyle bir programa katılmayan öğrencilere göre anlamlı bir şekilde daha yüksektir.

**Alternatif Hipotezi**  $H_1$  : X programına katılan öğrencilerin iletişim becerileri, böyle bir programa katılmayan öğrencilere göre anlamlı bir şekilde daha yüksek değildir.

Mann Whitney-U analizleri için veriler `read.csv2()` fonksiyonu yardımıyla okutulup, `iletirim` nesnesi olarak kaydedilir. `str()` fonksiyonu yardımıyla nesnenin yapısı incelenir.

```
iletirim <- read.csv2("iletirim.csv", header=TRUE)
iletirim
```

```
##      grup iletirim
## 1      1      41
## 2      1      80
## 3      1      82
## 4      1      59
## 5      1      69
## 6      1      80
## 7      1      65
## 8      2      30
## 9      2      20
## 10     2      40
## 11     2      60
## 12     2      47
## 13     2      55
## 14     2      79
```

```
str(iletirim)
```

```
## 'data.frame':  14 obs. of  2 variables:
## $ grup      : int  1 1 1 1 1 1 1 2 2 2 ...
## $ iletirim: int  41 80 82 59 69 80 65 30 20 40 ...
```

Dikkat edilirse `iletirim` nesnesine ait `grup` ve `iletirim` değişkenlerinin her ikisi de tamsayı (integer) değerlerdir. `grup` değişkeni faktör değişkene dönüştürülmelidir. Bu sayede deney ve kontrol olmak üzere iki farklı grup oluşturulabilir.

```
iletirim$grup <- factor(iletirim$grup, c(1,2), labels=c("Deney","Kontrol"))
iletirim$IletirimSira <- rank(iletirim$iletirim)
```

```
str(iletirim)
```

```
## 'data.frame':  14 obs. of  3 variables:
## $ grup      : Factor w/ 2 levels "Deney","Kontrol": 1 1 1 1 1 1 1 2 2 2 ...
## $ iletirim   : int  41 80 82 59 69 80 65 30 20 40 ...
## $ IletirimSira: num  4 12.5 14 7 10 12.5 9 2 1 3 ...
```

## 2.3. Betimsel İstatistiklerin Hesaplanması

Bir sonraki adımda verinin yapısını incelemek adına çeşitli gösterimlerde bulunulacaktır. **iletisim** veri setindeki değişkenlerin gruplara göre betimsel istatistikleri *tapply()* ve *summary()* fonksiyonları ile gösterilmiştir.

```
tapply(iletisim$iletisim, iletisim$grup, summary)
```

```
## $Deney
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    41      62      69      68      80      82
##
## $Kontrol
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  20.00  35.00  47.00  47.29  57.50  79.00
```

```
tapply(iletisim$IletisimSira, iletisim$grup, summary)
```

```
## $Deney
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  4.000  8.000 10.000  9.857 12.500 14.000
##
## $Kontrol
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.000  2.500  5.000  5.143  7.000 11.000
```

```
tapply(iletisim$IletisimSira, iletisim$grup, sum)
```

```
##   Deney Kontrol
##    69      36
```

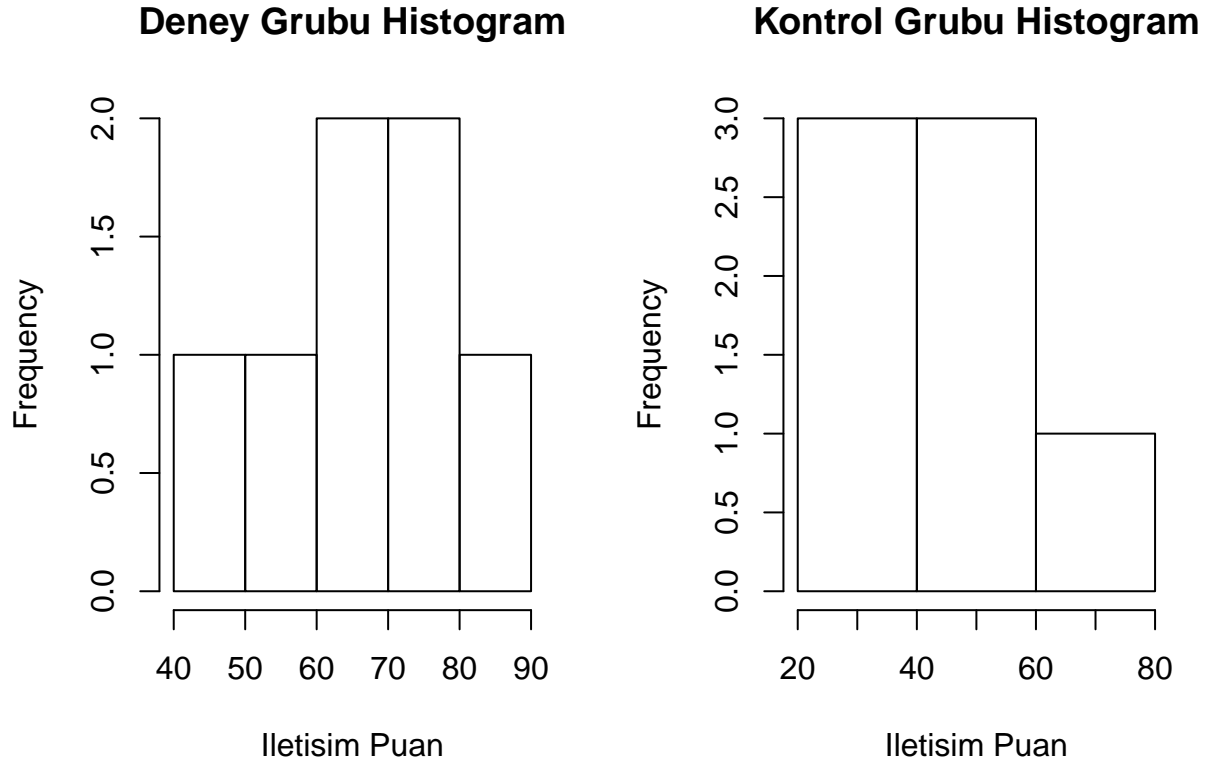
Yukarıda elde edilen çıktılar incelendiğinde üç özet değer elde edildiği görülebilir. İlki verilerin eşit aralık olarak kabul edildiği durumdaki değerlerdir. İkincisi ise verilerin sıralama düzeyinde olduğu duruma ait değerlerdir. İkinci çıktıdan Deney grubu için elde edilen 9,86 ile Kontrol grubu için elde edilen 5,14 değerleri Mann Whitney-U testinde rapor edilen Sıra Ortalamaları değerleridir. Son olarak ise Deney ve Kontrol grubuna ait Sıra Toplamları değerleri elde edilmiştir. Bu değer de Mann Whitney-U testinde rapor edilmesi istenen bir değerdir.

## 2.4. Varsayımların Test Edilmesi

Mann Whitney-U testinin varsayımları olan 1) iki grubun birbirinden bağımsız olması ve 2) bağımlı değişkenin en az sıralama ölçeğinde olması varsayımı test edilir. İkinci varsayım halihazırda sağlandığı için bu adımda birinci varsayıma ilişkin değerler gösterilecektir. Grupların ortancaları arasındaki farkı oraya koymak için bağımsız değişkenin her bir düzeyine göre bağımlı değişkenin dağılımının benzer olması beklenir. Dağılımların yerinin önemi yoktur yalnızca dağılımın şekillerinin birbirine benzemesi gerekmektedir.

Deney ve kontrol grubu grafiklerini aynı grafikte göstermek için grafik penceresi *par()* fonksiyonu ile belirlenir. Bu sayede 1 satır ve 2 sütunlu bir boş çerçeve oluşturulmuş olur. Daha sonra bağımsız değişkenin her düzeyleri olan Deney ve Kontrol grupları için histogramlar aşağıdaki gibi çizilir.

```
par(mfrow=c(1,2))
hist(iletisim$iletisim[iletisim$grup == "Deney"],
     main = "Deney Grubu Histogram", xlab= "İletişim Puan")
hist(iletisim$iletisim[iletisim$grup == "Kontrol"],
     main = "Kontrol Grubu Histogram", xlab= "İletişim Puan")
```



```
par(mfrow=c(1,1))
```

Deney ve kontrol grubuna ait histogramların şekilleri birbirinden farklı çıkmıştır. Bu bakımdan ortanca değerler yerine ortalama sıra (mean rank) değerleri rapor edilmelidir. Eğer histogramlar birbirine benzer çıkmış olsaydı iletişim becerilerini farklarını açıklamak için ortanca değerler kullanılabilecekti.

## 2.5. Mann Whitney-U testinin oluşturulması

Mann Whitney-U testi sıfır hipotezini test etmektedir.  $p$  değeri belirlenen kritik değerden (0,05) küçük olduğu durumda sıfır hipotezi reddedilir. `wilcox.test(sayısal_değişken ~ faktör_değişken)*` formu kullanılarak hesaplamalar yapılır.

```
wilcox.test(iletisim$Iletisim ~ iletisim$grup, alternative = "greater")
```

```
## Warning in wilcox.test.default(x = c(4, 12.5, 14, 7, 10, 12.5, 9), y =
```

```
## c(2, : cannot compute exact p-value with ties
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: iletisim$Iletisim by iletisim$grup
```

```
## W = 41, p-value = 0.02035
```

```
## alternative hypothesis: true location shift is greater than 0
```

Analiz sonucunda  $W=41$  değeri ve  $p=0,02035$  olmak üzere iki değer elde edilir. Wilcoxon W değeri sıra değerlerinin toplamının en düşüğünü gösterir ve  $p$  değerini hesaplamak için kullanılır.

## 2.6. Sonuçların Raporlanması

Mann Whitney-U testi sonuçlarına göre anti-sosyal davranışları gözlenen çocuklardan X programına katılanlarla, böyle bir programa katılmayanların uygulama sonrasındaki iletişim becerileri Tablo 2.2’de verilmiştir. Buna göre, dört haftalık bir deneysel çalışma sonunda, X programına katılan çocuklar ile böyle bir programa katılmayan çocukların iletişim becerileri arasında anlamlı bir fark olduğu bulunmuştur ( $W=41$ ,  $p<0,05$ ). Sıra ortalamaları dikkate alındığında, X programına katılan çocukların, programa katılmayan çocuklara göre iletişim becerilerinin daha yüksek olduğu anlaşılmaktadır. Bu bulgu, X programının, anti-sosyal davranış gösteren çocukların iletişim becerilerini arttırmada etkili olduğunu gösterir.

**Tablo 2.2** İletişim Becerilerinin Gruba Göre Mann Whitney-U Testi Sonuçları

Grup	n	Sıra Ortalaması	Sıra Toplamı	W	p
Deney	7	9,86	69	41,0	0,02
Kontrol	7	5,14	36		

## 3. Wilcoxon İşaretli Sıralar Testi

İlişkili iki örnekleme ait puanlar arasındaki farkın anlamlılığını test etmek için kullanılır. Parametrik testlerdeki **bağımlı örneklem t-testi** ile benzerdir. Wilcoxon İşaretli Sıralar testinde gözlem çiftleri birbirinden bağımsızdır ve bağımlı değişkenler sıralama düzeyindedir.

Wilcoxon İşaretli Sıralar testinde analizler fark puanları üzerinden yürütülür. Fark puanları küçükten büyüğe doğru, 1’den başlayarak işareti dikkat edilmeksizin sıra sayıları verilmesi temeline dayanır. Daha sonra + ve - işaretli olan fark puanlarının sıra sayıları toplanır. Test edilen durum gerçekte, bu iki sıra sayılarının toplamı arasındaki farktır. Aynı puanlar analiz dışı tutulur ve analiz küçük olan sıra toplamı üzerine kurulur. Z istatistiği kullanılır.

Wilcoxon İşaretli Sıralar testi az denekli yürütülen grup içi karşılaştırmalarda kullanılır. Deneklerin fark puanlarının normal dağılım göstermediği durumlarda bağımlı örneklem t-testi yerine tercih edilir.

Wilcoxon İşaretli Sıralar analizleri *wilcox.test()* fonksiyonu ile yürütülebilir. Fonksiyonunun kullanımı aşağıdaki gibidir.

```
wilcox.test(x, y = NULL,
            alternative = c("two.sided", "less", "greater"),
            paired= TRUE, conf.int = FALSE, conf.level = 0.95, ...)
```

Daha önce Tablo 2.1’de açıklanan *wilcox.test()* fonksiyonunun argümanlarından “paired=TRUE” argümanı eklenir.

### 3.1. Testin Aşamaları

#### Örnek

Konuşma becerilerini geliştirmede kullanılmak amacıyla hazırlanan X programının etkililiği araştırılıyor. Okulöncesi eğitime devam eden ve konuşma güçlüğü çeken sekiz çocuğun uygulama öncesi ve sonrası bir konuşma testinden aldıkları puanlar “**konusma.csv**” dosyasına kaydedilmiştir.

### 3.2. Araştırma Problemi ve Hipotezlerin Oluşturulması

**Araştırma Problemi** Çocukların konuşma becerileri, uygulanan X programı sonrasında anlamlı bir farklı mıdır?

**Sıfır Hipotezi**  $H_0$  : Çocukların X programına ait öncesi ve sonrası konuşma becerileri arasında anlamlı bir farkdır.

**Alternatif Hipotezi**  $H_1$  : Çocukların X programına ait öncesi ve sonrası konuşma becerileri arasında anlamlı bir fark yoktur.

Wilcoxon İşaretli Sıralar analizleri için veriler `read.csv2()` fonksiyonu yardımıyla okutulup, *iletisim* nesnesi olarak kaydedilir. `str()` fonksiyonu yardımıyla nesnenin yapısı incelenir.

```
konusma <- read.csv2("konusma.csv", header=TRUE)
konusma
```

```
##   ontest sontest izleme
## 1      2      5      4
## 2      6      6      6
## 3      4      6      5
## 4      5      3      4
## 5      3      8      7
## 6      1      4      4
## 7      4      7      6
## 8      2      4      4
```

```
str(konusma)
```

```
## 'data.frame':   8 obs. of  3 variables:
## $ ontest : num  2 6 4 5 3 1 4 2
## $ sontest: num  5 6 6 3 8 4 7 4
## $ izleme : num  4 6 5 4 7 4 6 4
```

Dikkat edilirse **konusma** nesnesine faktör değişken bulunmamaktadır. Burada aynı bireye ait öntest ve sontest puanları ilişkili ölçümleri belirtmektedir. İlişkili ölçümlerde fark değerleri üzerinden işlemler yapıldığından sontest puanlarından öntest puanlarını çıkararak fark değişkeni tanımlanmış ve daha sonra sıra numarası verilmiştir. Ayrıca `cut()` fonksiyonu kullanılarak sıfırdan düşük değerlere “Negatif Sıra”, sıfırdan büyük değerlere ise “Pozitif Sıra”, sıfıra eşit değerlere de “Eşit” ismi verilerek üç faktörlü grup değişkeni oluşturulmuştur.

```
konusma$fark <- konusma$sontest - konusma$ontest

konusma$grup <- cut(konusma$fark, c(-10,-0.99, 0.099,+10),
                    labels = c('Negatif Sıra','Eşit','Pozitif Sıra'), right = FALSE)

konusma$farkSira <- rank(abs(konusma$fark))
str(konusma)
```

```
## 'data.frame':   8 obs. of  6 variables:
## $ ontest : num  2 6 4 5 3 1 4 2
## $ sontest: num  5 6 6 3 8 4 7 4
## $ izleme : num  4 6 5 4 7 4 6 4
## $ fark   : num  3 0 2 -2 5 3 3 2
## $ grup   : Factor w/ 3 levels "Negatif Sıra",...: 3 2 3 1 3 3 3 3
## $ farkSira: num  6 1 3 3 8 6 6 3
```

### 3.3. Betimsel İstatistiklerin Hesaplanması

Bir sonraki adımda verinin yapısını incelemek adına çeşitli gösterimlerde bulunulacaktır. **konusma** veri setindeki değişkenlerin gruplara göre betimsel istatistikleri `tapply()` ve `summary()` fonksiyonları ile gösterilmiştir.

```
summary(konusma)
```

```
##      ontest      sontest      izleme      fark
## Min.   :1.000   Min.   :3.000   Min.   :4.0   Min.   : -2.0
## 1st Qu.:2.000   1st Qu.:4.000   1st Qu.:4.0   1st Qu.: 1.5
## Median :3.500   Median :5.500   Median :4.5   Median : 2.5
## Mean   :3.375   Mean   :5.375   Mean   :5.0   Mean   : 2.0
## 3rd Qu.:4.250   3rd Qu.:6.250   3rd Qu.:6.0   3rd Qu.: 3.0
## Max.   :6.000   Max.   :8.000   Max.   :7.0   Max.   : 5.0
##      grup      farkSira
## Negatif Sıra:1   Min.   :1.0
## Eşit          :1   1st Qu.:3.0
## Pozitif Sıra:6   Median :4.5
##                Mean   :4.5
##                3rd Qu.:6.0
##                Max.   :8.0
```

```
tapply(konusma$farkSira, konusma$grup, summary)
```

```
## $`Negatif Sıra`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##        3      3      3      3      3      3
##
## $Eşit
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##        1      1      1      1      1      1
##
## $`Pozitif Sıra`
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.000  3.750  6.000  5.333  6.000  8.000
```

```
tapply(konusma$farkSira, konusma$grup, sum)
```

```
## Negatif Sıra      Eşit Pozitif Sıra
##           3           1           32
```

Yukarıda elde edilen çıktılar incelendiğinde üç özet değer elde edildiği görülebilir. İlk verilerin eşit aralık olarak kabul edildiği durumdaki değerlerdir. İkincisi ise verilerin sıralama düzeyinde olduğu duruma ait değerlerdir. İkinci çıktıdan Sıra Ortalamaları değerleri elde edilir. Son olarak ise Sıra Toplamları değerleri elde edilmiştir. Elde edilen değerler Wilcoxon İşaretli Sıralar testinde rapor edilir.

### 3.4. Varsayımların Test Edilmesi

Wilcoxon İşaretli Sıralar testinin fark dağılımlarının normal olup olmadığına bakılır. Bunun için fark değerlerinin histogramı çizilmiştir.

```
median(konusma$sontest)
```

```
## [1] 3.5
```

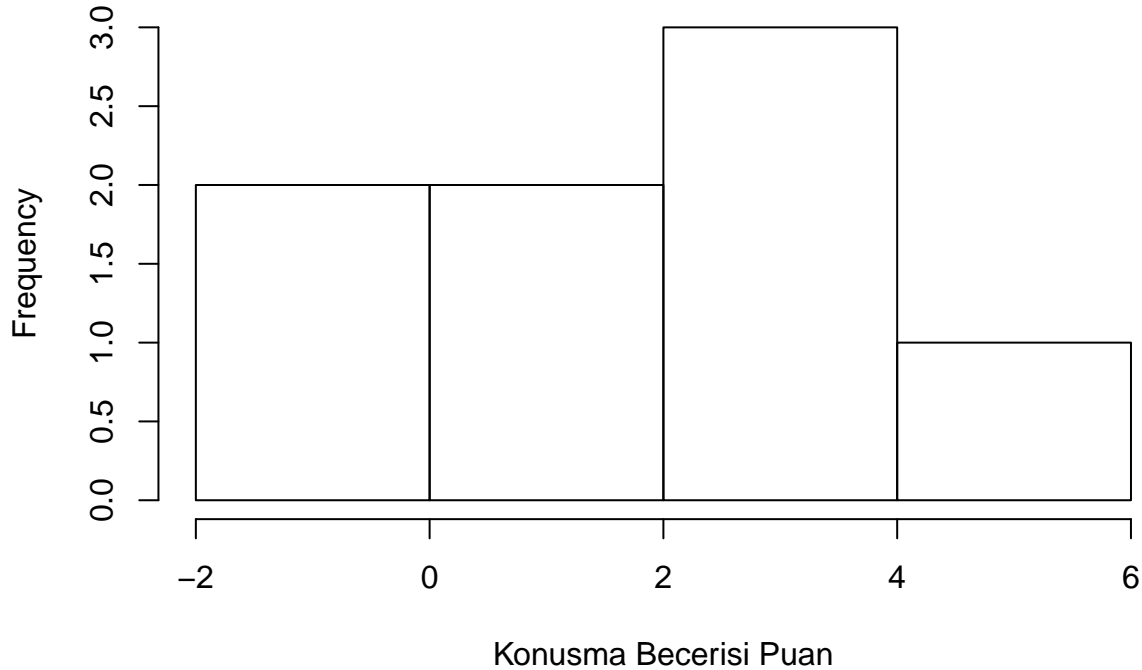
```
median(konusma$sontest)
```

```
## [1] 5.5
```

```
hist(konusma$fark, main = "Fark Ortalamaları Histogramı", xlab= "Konuşma Becerisi Puan")
```



## Fark Ortalamalari Histogrami



Grafik incelendiğinde eşleşmiş ölçümlerin farklarının normal olmadığı ve böylece parametrik testlerin yerine nonparametrik testlerin uygulanması gerektiği yorumu yapılabilir. Parametrik testlerde daima ortalama yerine ortanca değerlerinin yorumlanmasına dikkat ediniz.

### 3.5. Wilcoxon İşaretli Sıralar testinin oluşturulması

Wilcoxon İşaretli Sıralar testi sıfır hipotezini test etmektedir.  $p$  değeri belirlenen kritik değerden (0,05) küçük olduğu durumda sıfır hipotezi reddedilir. `wilcox.test(sayısal_değişken, sayısal_değişken, paired=TRUE)` formu kullanılarak hesaplamalar yapılır.

```
wilcox.test(konusma$ontest, konusma$sontest, paired = TRUE)
```

```
## Warning in wilcox.test.default(konusma$ontest, konusma$sontest, paired =  
## TRUE): cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(konusma$ontest, konusma$sontest, paired =  
## TRUE): cannot compute exact p-value with zeroes
```

```
##
```

```
## Wilcoxon signed rank test with continuity correction
```

```
##
```

```
## data: konusma$ontest and konusma$sontest
```

```
## V = 2, p-value = 0.04858
```

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

Analiz sonucunda  $V=2$  değeri ve  $p=0,0485$  olmak üzere iki değer elde edilir. Wilcoxon  $W$  değeri sıra değerlerinin toplamının en düşüğünü gösterir ve  $p$  değerini hesaplamak için kullanılır.  $V$  değeri hesaplanan  $Z$  değerini belirtmektedir.

### 3.6. Sonuçların Raporlanması

sonuçlarına göre çocukların öncesi ve sonrası konuşma becerilerinin anlamlı bir farklılık gösterip göstermediğine ilişkin Wilcoxon İşaretli Sıralar testi sonuçları Tablo 3.1’de verilmiştir. Analiz sonuçları, araştırmaya katılan çocukların konuşma becerileri testinden aldıkları deney öncesi ve sonrası puanları arasında anlamlı bir fark olduğunu göstermektedir ( $Z=2$ ;  $p=<0,05$ ). Fark puanlarının sıra ortalaması ve toplamı dikkate alındığında, gözlenen bu farkın pozitif sıralar, yani son test puanı lehine olduğu görülmektedir. Bu sonuçlara göre, düzenlenen X programının çocukların konuşma becerilerini geliştirmede önemli bir etkisinin olduğu söylenebilir.

**Tablo 3.1** İletişim Becerilerinin Gruba Göre Mann Whitney-U Testi Sonuçları

Sontest-Öntest	n	Sıra Ortalaması	Sıra Toplamı	Z	p
Negatif Sıra	1	3	3	2	0,0485
Pozitif Sıra	6	5,33	32		
Eşit	1				

## 4. Kruskal Wallis Testi

Kruskal Wallis testi Mann Whitney-U testinin ikiden fazla örnekleme uygulanan halidir. Analizde k tane örneklemin bağımlı değişken üzerinde anlamlı farklılık gösterip göstermediği incelenir. ANOVA testinin nonparametrik halidir ve varsayımları arasında a) bağımlı değişkenin en az sıralama ölçeği düzeyinde olması b) gözlemlerin birbirinden bağımsız olmasını gerektirir. ANOVA’dan farklı olarak her bir alt gruba ait normal dağılım ve varyansların eşitliği varsayımı bulunmaz.

Kruskal Wallis testinde hipotezler ortanca değerler üzerine kurulmaz. Aksine hipotezler bir grupta gözlemlerin diğer gruptan büyük/eşit/küçük olduğu üstüne kurulmaktadır.

Kruskal Wallis analizleri `kruskal.test()` fonksiyonu ile yürütülebilir. Fonksiyonunun kullanımı aşağıdaki gibidir.

```
kruskal.test(x, g, ...)
```

### 4.1. Testin Aşamaları

#### Örnek

Okuma güçlüğü çeken bir grup öğrencinin, okuma becerilerini geliştirmede üç farklı eğitim programının (A-B-C) etkililikleri araştırılıyor. Bu amaçla, ilkokul birinci sınıf öğrencileri içinden belirlenen 12 öğrenci, yansız olarak 3 gruba atanıyor ve öğrenciler, normal eğitimlerinin yanı sıra üç haftalık bir okuma becerilerini geliştirme eğitimine katılıyorlar. Çalışmaya ilişkin denemelik veriler **okuma.csv** isimli dosyada bulunmaktadır.

### 4.2. Araştırma Problemi ve Hipotezlerin Oluşturulması

**Araştırma Problemi** Öğrencilerin okuma becerileri, alınan eğitim programına göre anlamlı farklılık göstermekte midir?

**Sıfır Hipotezi**  $H_0$  : Öğrencilerin okuma becerileri, alınan eğitim programına göre anlamlı farklılık göstermemektedir.

**Alternatif Hipotezi**  $H_1$  : Öğrencilerin okuma becerileri, alınan eğitim programına göre anlamlı farklılık göstermemektedir.

Kruskal Wallis analizleri için veriler `read.csv2()` fonksiyonu yardımıyla okutulup, *okuma* nesnesi olarak kaydedilir. `str()` fonksiyonu yardımıyla nesnenin yapısı incelenir.

```
okuma <- read.csv2("okuma.csv", header=TRUE)
okuma
```

```
##      yontem okuma
## 1      A      9
## 2      A     12
## 3      A     10
## 4      A      6
## 5      B     24
## 6      B     15
## 7      B     18
## 8      B     14
## 9      C     22
## 10     C     29
## 11     C     33
## 12     C     27
```

```
str(okuma)
```

```
## 'data.frame':   12 obs. of  2 variables:
## $ yontem: Factor w/ 3 levels "A","B","C": 1 1 1 1 2 2 2 2 3 3 ...
## $ okuma : num  9 12 10 6 24 15 18 14 22 29 ...
```

Dikkat edilirse **okuma** nesnesinde *yontem* faktör değişken iken *okuma* ise sayısal değişkendir. Veri setinde okuma puanlarına göre sıralamalar `rank()` fonksiyonu ile oluşturulup *okumaFark* değişkeni olarak kaydedilmiştir.

```
okuma$okumaFark <- rank(okuma$okuma)
```

```
okuma
```

```
##      yontem okuma okumaFark
## 1      A      9          2
## 2      A     12          4
## 3      A     10          3
## 4      A      6          1
## 5      B     24          9
## 6      B     15          6
## 7      B     18          7
## 8      B     14          5
## 9      C     22          8
## 10     C     29         11
## 11     C     33         12
## 12     C     27         10
```

### 4.3. Betimsel İstatistiklerin Hesaplanması

Bir sonraki adımda verinin yapısını incelemek adına çeşitli gösterimlerde bulunulacaktır. Okuma puanlarının yöntemlere göre betimsel istatistikleri `tapply()` ve `summary()` fonksiyonları ile gösterilmiştir.

```
summary(okuma)
```

```
##      yontem      okuma      okumaFark
## A:4      Min.       : 6.00      Min.       : 1.00
```

```
## B:4 1st Qu.:11.50 1st Qu.: 3.75
## C:4 Median :16.50 Median : 6.50
## Mean :18.25 Mean : 6.50
## 3rd Qu.:24.75 3rd Qu.: 9.25
## Max. :33.00 Max. :12.00
```

```
tapply(okuma$okumaFark, okuma$yontem, summary)
```

```
## $A
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 1.00 1.75 2.50 2.50 3.25 4.00
##
## $B
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 5.00 5.75 6.50 6.75 7.50 9.00
##
## $C
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 8.00 9.50 10.50 10.25 11.25 12.00
```

```
tapply(okuma$okumaFark, okuma$yontem, sum)
```

```
## A B C
## 10 27 41
```

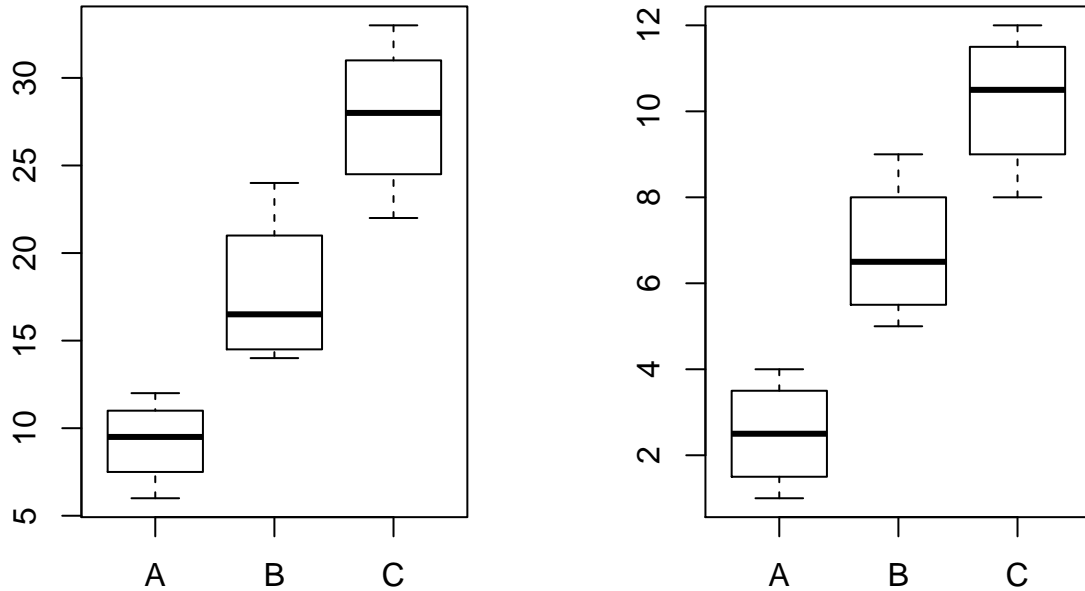
Yukarıda elde edilen çıktılar incelendiğinde üç özet değer elde edildiği görülebilir. İlk verilerin eşit aralık olarak kabul edildiği durumdaki değerlerdir. İkincisi ise verilerin sıralama düzeyinde olduğu duruma ait değerlerdir. Burada her bir yöntem için elde edilen ortalama değerleri A yöntemi için 2,50; B yöntemi için 6,75 ve C yöntemi için 10,25 olarak bulunmuştur. Bu değerler rapor edilmesi gereken Sıra ortalaması değerleridir. Son olarak ise A yöntemi için 10; B yöntemi için 27 ve C yöntemi için 41 olarak Sıra Toplamları değerleri elde edilmiştir. Elde edilen değerler Kruskal Wallis testinde rapor edilir. Elde edilen sonuçlar incelendiğinde C yönteminin öğrencilerin okuma becerilerine diğer iki yöntemden daha fazla etki ettiği görülmüştür.

#### 4.4. Varsayımların Test Edilmesi

Kruskal Wallis testinde ortanca değerlerin nasıl farklılaştığını belirlemek adına her üç yöntem için kutu grafikleri çizilir. Kutu grafiğinin çiziminde *boxplot(bağımlıdeğişken~bağımsızdeğişken)* formatı kullanılır.

```
par(mfrow=c(1,2))
boxplot(okuma$okuma~okuma$yontem, main= "Yöntemlere Göre Okuma Becerisi Puanları")
boxplot(okuma$okumaFark~okuma$yontem, main= "Yöntemlere Göre Okuma Becerisi Sırası")
```

## öntemlere Göre Okuma Becerisi PuYöntemlere Göre Okuma Becerisi S



```
par(mfrow=c(1,1))
```

Kutu grafiğinde kutuların içindeki kalın yatay çizgiler ortanca değerleri ifade eder. Kutu grafiğinin alt ve üst değerleri de dağılımın en düşük ve en yüksek noktalarını belirtir. Burada da görüldüğü gibi C yönteminin diğer iki yöntemden daha fazla okuma becerisine katkı sağladığı görülmektedir. Sağdaki grafikte ise yöntemlere göre okuma becerilerinden elde edilen sıra değerleri grafiği çizilmiştir. Her iki grafik birbiriyle tutarlılık göstermektedir.

Grafiklerin daha renkli gösterimleri için aşağıdaki kodları kullanabilirsiniz.

```
library("ggpubr")
```

```
## Loading required package: ggplot2
```

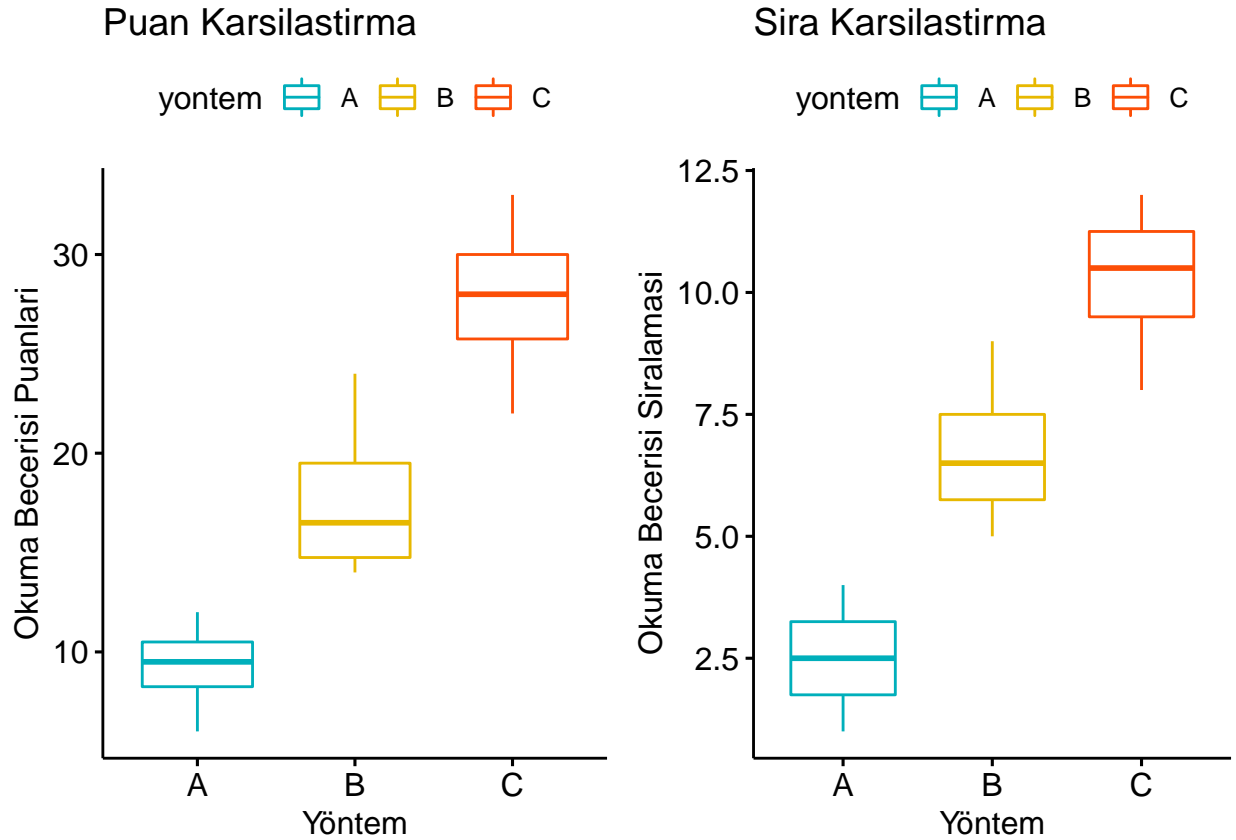
```
## Loading required package: magrittr
```

```
library("ggplot2")
```

```
g1 <- ggboxplot(okuma, x = "yontem", y = "okuma",  
  color = "yontem", palette = c("#00AFBB", "#E7B800", "#FC4E07"),  
  order = c("A", "B", "C"), main="Puan Karşılaştırma",  
  ylab = "Okuma Becerisi Puanları", xlab = "Yöntem")
```

```
g2 <- ggboxplot(okuma, x = "yontem", y = "okumaFark",  
  color = "yontem", palette = c("#00AFBB", "#E7B800", "#FC4E07"),  
  order = c("A", "B", "C"), main="Sıra Karşılaştırma",  
  ylab = "Okuma Becerisi Sıralaması", xlab = "Yöntem")
```

```
ggarrange(g1, g2, ncol = 2, nrow = 1)
```



#### 4.5. Kruskal Wallis testinin oluşturulması

Kruskal Wallis testinin amacı sıfır hipotezini test etmektedir.  $p$  değeri belirlenen kritik değerden (0,05) küçük olduğu durumda sıfır hipotezi reddedilir. `kruskal.test(bağımlı_değişken ~ bağımsız_değişken)` formu kullanılarak hesaplamalar yapılır.

```
kruskal.test(okuma$okuma ~ okuma$yontem)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: okuma$okuma by okuma$yontem
## Kruskal-Wallis chi-squared = 9.2692, df = 2, p-value = 0.00971
```

Analiz sonucunda  $sd=(k-1)$  üzerinden kestirilen ki-kare istatistiği değeri hesaplanır.  $x^2=9,26$  değeri ve  $p=0,001$  olmak üzere iki değer elde edilir.  $p$  değerinin kritik değer olan 0,05'ten küçük olması nedeniyle yöntemler arasında anlamlı bir farkın olduğu yorumu yapılabilir. ,

Farkların hangi gruplar arasında olduğunu belirlemek için ANOVA analizlerindeki gibi Kruskal Wallis testinde de post-hoc testlerine başvurulur. Kruskal Wallis için post-hoc testi `pairwise.wilcox.test()` fonksiyonu yardımı ile yapılmaktadır. Fonksiyonunun kullanımı aşağıdaki gibidir:

```
pairwise.wilcox.test(x, g, p.adjust.method = p.adjust.methods, ...)
```

`pairwise.wilcox.test()` fonksiyonunun argümanları Tablo 4.1'de açıklanmıştır.

**Tablo 4.1** *pairwise.wilcox.test()* fonksiyonu argümanları ve açıklamaları

Argümanlar	Açıklamalar
x	Bağımlı Değişken
g	Kategorik bağımsız değişken
p.adjust.method	Nonparametrik ikili karşılaştırma yöntemleri:“holm”,“bonferroni”,“fdr”,“BY” ve “none”

```
pairwise.wilcox.test(okuma$okuma, okuma$yontem, p.adj='fdr')
```

```
##  
## Pairwise comparisons using Wilcoxon rank sum test  
##  
## data: okuma$okuma and okuma$yontem  
##  
## A B  
## B 0.043 -  
## C 0.043 0.057  
##  
## P value adjustment method: fdr
```

Post-hoc analizi sonucu p değerlerin karşılaştırdığı bir matris elde edilir. Bu değerlere bakılarak  $p < 0,05$  koşulunu sağlayan ikililer belirlenir. Burada A-B ve A-C grupları arasında anlamlı bir fark olduğu; B-C grupları arasında anlamlı fark bulunmadığı yorumu yapılabilir.

#### 4.6. Sonuçların Raporlanması

Kruskal Wallis testi sonuçlarına göre farklı öğretim yöntemine göre eğitim alan, okuma güçlüğü çeken çocukların program sonrasındaki okuma becerileri testinden aldıkları Kruskal Wallis testi sonuçları Tablo 4.2’de verilmiştir.

Analiz sonuçları, deneysel çalışmaya katılan çocukların okuma becerileri testinden aldıkları puanların, aldıkları öğretim yöntemine göre anlamlı bir şekilde farklılaştığını göstermektedir ( $x^2(sd=2, n=12) = 9,26; p<0,001$ ). Bu bulgu, kullanılan üç yöntemin çocukların okuma becerilerini artırmada farklı etkilere sahip olduğunu gösterir. Grupların sıra ortalamaları dikkate alındığında, uygulama sonrasında en yüksek okuma becerisi sırasının C programına katılanların sahip olduğu, bunu B ve A’nın izlediği görülmektedir.

Gruplar arasından gözlenen anlamlı farkın hangi gruplar arasında çıktığını belirlemek adına Wilcoxon ikili karşılaştırma testi (post-hoc) kullanılmıştır. Elde edilen sonuçlara bakıldığında A-B ve A-C grupları arasında (B ve C lehine) anlamlı bir fark olduğu; B-C grupları arasında anlamlı fark bulunmadığı yorumu yapılabilir.

**Tablo 4.2** *Okuma Becerilerinin Uygulanan Yönteme Göre Kruskal Wallis Testi Sonuçları*

Yöntem	n	Sıra Ortalaması	sd	ki-kare	p	Anlamlı Fark
A	4	2,50	2	9,26	0,01	A-B
B	4	6,75				A-C
C	4	10,25				

#### 5. Ki-kare Testi

Ki-kare testi kategorik bir değişkenin düzeylerine giren birey ya da nesnelerin anlamlı bir farklılık gösterip göstermediğini test eder. İyi uyum testi olarak da bilinir. Kategorik değişkenler arasındaki farkın anlam-

lıgını inceler. Bu bakımdan gözlenen değerler ile beklenen değerlerin birbirinden anlamlı olarak farklılaşıp farklılaşmadığına bakar.

Tek örneklem ve iki örneklem üzerinden ki-kare testleri yapılabilmektedir. Bu bölümde her iki tür de anlatılacaktır.

Ki-kare testinde her bir kategoriye ait gözlenen ve beklenen değerler arasındaki farkın karesinin beklenen değere bölünmesiyle elde edilen bir ki-kare istatistiği değeri elde edilir ve her koşul için elde edilen bu değerler toplanır. Beklenen ve gözlenen değerler arasındaki fark azaldıkça yani birimlerin kategorilere dağılımının uyum düzeyi arttıkça ki-kare istatistiğinin anlamlı çıkma olasılığı düşecektir (yani farklılık olmayacaktır). Bu testin kullanılabilmesi için beklenen değerlerin beşten küçük kategori sayısının, toplam kategori sayısının %20'sini aşmaması ve tüm kategorilerde bu değerlerin 1'den büyük olması gerekir. Bu koşul sağlanmıyor ise kategorilerde birleştirme yapılması önerilir.

Tek örneklem ki-kare testi tek grup araştırmalarında tek bir değişkene ilişkin değişkenliğin incelendiği deneysel ya da tarama çalışmalarında kullanılabilir.

Ki-kare analizleri `chisq.test()` fonksiyonu ile yürütülebilir. Fonksiyonunun ve argümanların kullanımı Tablo 5.1'de gösterilmiştir.

```
chisq.test(x, y = NULL, correct = TRUE, p = rep(1/length(x)))
```

**Tablo 5.1** `chisq.test()` fonksiyonu argümanları ve açıklamaları

Argümanlar	Açıklamalar
x	Sayısal vektör ya da matris
y	İkinci değişkene ait sayısal vektör ya da matris
correct	Yates süreklilik düzeltilmesi kullanımı
p	Beklenen olasılık dağılımı

## 5.1. Testin Aşamaları

### Örnek

Anaokulundaki çocukların tercih ettikleri oyun türünü gözleyen araştırmacı gözlem sonuçlarına göre çocukların oynadıkları oyunlar arasında bir uyum olup olmadığını incelemek istemektedir.

## 5.2. Araştırma Problemi ve Hipotezlerin Oluşturulması

**Araştırma Problemi** Çocukların tercih ettikleri oyun türleri arasında anlamlı bir farklılık var mıdır?

**Sıfır Hipotezi**  $H_0$  : Çocukların tercih ettikleri oyun türleri arasında anlamlı bir farklılık vardır.

**Alternatif Hipotezi**  $H_1$  : Çocukların tercih ettikleri oyun türleri arasında anlamlı bir farklılık yoktur.

Yapılan araştırmada çocukların oynadıkları oyunların üç farklı kategoride toplandığı belirlenmiştir. Kategoriler ilişkin gözlem sayıları aşağıdaki gibi belirtilmiştir. Fiziksel: 11 Küp oyunları: 7 Dramatik oyunları:2

```
oyun <- c(11,7,2)
oyun
```

```
## [1] 11 7 2
```

```
str(oyun)
```

```
## num [1:3] 11 7 2
```



### 5.3. Betimsel İstatistiklerin Hesaplanması

Ki-kare analizlerinde veriye ait betimsel istatistiklerden sadece yüzdelik oranlara ilişkin ifadeler elde edilebilir. `prop.table()` fonksiyonu ile her bir oyun türüne ait oranlar belirlenir.

```
prop.table(oyun)
```

```
## [1] 0.55 0.35 0.10
```

Sonuçlar incelendiğinde tüm oyunlar içinde fiziksel oyunların %55, küp oyunlarının %35 ve dramatik oyunların ise %10'luk bir paya sahip olduğunu görmekteyiz. Bu elde edilenler gözlenen değerlerdir. Beklenen değerler ise her bir oyuna 1/3 olacak şekilde dağılımıdır. Bu varsayımlar ile ki-kare analizi aşağıdaki gibi formüle edilir.

### 5.4. Ki-kare testinin oluşturulması

Ki-kare testinin kategoriler arasındaki anlamlı farkları test etmektir.  $p$  değeri belirlenen kritik değerden (0,05) küçük olduğu durumda sıfır hipotezi reddedilir. `chisq.test(sayısal_değişken, p=beklenen_değer)` formu kullanılarak hesaplamalar yapılır.

```
oyun_sonuc <- chisq.test(oyun, p = c(1/3, 1/3, 1/3))
oyun_sonuc
```

```
##
## Chi-squared test for given probabilities
##
## data: oyun
## X-squared = 6.1, df = 2, p-value = 0.04736
```

```
oyun_sonuc$expected
```

```
## [1] 6.666667 6.666667 6.666667
```

```
oyun_sonuc$observed
```

```
## [1] 11 7 2
```

```
oyun_sonuc$statistic
```

```
## X-squared
##      6.1
```

```
oyun_sonuc$p.value
```

```
## [1] 0.04735892
```

### 5.5. Sonuçların Raporlanması

Analiz sonucunda tek örneklem ki-kare testi sonucuna göre gözlem yapılan 11'i fiziksel, 7'si küp ve 2'si dramatik türden oyun oynanmıştır. En az tercih edilen oyun türü dramatik en çok ise fizikseldir. Çocukların tercih ettikleri oyun türleri arasında gözlenen fark anlamlı bulunmuştur ( $X^2=6,1$ ;  $p < 0,05$ )

**Tablo 5.2** Tercih edilen oyun türüne göre oynama sıklıklarının karşılaştırılmasına ilişkin  $X^2$  testi sonuçları

Gruplar	Gözlenen	Beklenen	$X^2$	sd	p
Fiziksel	11	6,7	6,1	2	0,047
Küp	7	6,7			

Gruplar	Gözlenen	Beklenen	$X^2$	sd	p
Dramatik	2	6,7			
Toplam	20				