İstatistiksel Yöntemler 2 Ödevi

Ramazan Erduran – 21821809

İçindekiler Tablosu

[Tek Yönlü ANOVA: 1](#_Toc91253978)

[Normallik testi: 1](#_Toc91253979)

[Varyans homojenliği: 1](#_Toc91253980)

[ANOVA hesabı: 2](#_Toc91253981)

[Post-Hoc testi: 2](#_Toc91253982)

[Özetleyici istatistikler: 2](#_Toc91253983)

[Tekrarlı Ölçümlü ANOVA: 3](#_Toc91253984)

[Verileri manipüle etme: 4](#_Toc91253985)

[Ortalama ve St. Sapma değerleri: 5](#_Toc91253986)

[Extreme Aykırı değer kontrolü: 5](#_Toc91253987)

[Normallik testi: 5](#_Toc91253988)

[Varyans homojenliği: 6](#_Toc91253989)

[ANOVA testi: 6](#_Toc91253990)

[Rastgele Blok: 6](#_Toc91253991)

[Normallik testi: 7](#_Toc91253992)

[Verileri manipüle etme: 7](#_Toc91253993)

[Varyans homojenliği: 7](#_Toc91253994)

[ANOVA testi: 8](#_Toc91253995)

[Post Hoc testi: 8](#_Toc91253996)

[Latin Kare: 8](#_Toc91253997)

[Veriyi manipüle etme: 9](#_Toc91253998)

[ANOVA testi: 9](#_Toc91253999)

[Ortalama karşılaştırma testi: 10](#_Toc91254000)

[Çok Etkenli Deney Düzeni: 11](#_Toc91254001)

[Verileri manipüle etme: 12](#_Toc91254002)

[ANOVA testi: 12](#_Toc91254003)

[Etkileşim grafikleri: 12](#_Toc91254004)

[Kaynakça: 13](#_Toc91254005)

## Tek Yönlü ANOVA:

Verileri içeri aktarma.

library(readxl)  
crop\_data <- read\_excel("C:/Users/lenovo/Documents/GitHub/StatMeth2/Statistical-Methods-2/One\_way\_ANOVA/one\_way\_data.xlsx")  
head(crop\_data,3)

## # A tibble: 3 x 2  
## fertilizer yield  
## <dbl> <dbl>  
## 1 1 177.  
## 2 1 178.  
## 3 1 176.

### Normallik testi:

H0: Verilerin dağılımı ile normal dağılım arasında fark yoktur.  
Hs: Verilerin dağılımı ile normal dağılım arasında fark vardır.

shapiro.test(crop\_data$yield)

##   
## Shapiro-Wilk normality test  
##   
## data: crop\_data$yield  
## W = 0.989, p-value = 0.6135

%95 güven düzeyinde H0 reddedilemez. Verilerin dağılımı ile normal dağılım arasında fark yoktur.

### Varyans homojenliği:

H0: 1 = 2  
Hs: 1 2

bartlett.test(yield~fertilizer,data=crop\_data)

##   
## Bartlett test of homogeneity of variances  
##   
## data: yield by fertilizer  
## Bartlett's K-squared = 1.0622, df = 2, p-value = 0.5879

%95 güven düzeyinde H0 reddedilemez. Varyanslar homojendir.

### ANOVA hesabı:

H0: Gübreler arasında fark yoktur.  
Hs: Gübreler arasında fark vardır.

crop\_data$fertilizer <- as.factor(crop\_data$fertilizer)  
rez\_aov=lm( crop\_data$yield ~ crop\_data$fertilizer)   
anova(aov(rez\_aov))

## Analysis of Variance Table  
##   
## Response: crop\_data$yield  
## Df Sum Sq Mean Sq F value Pr(>F)   
## crop\_data$fertilizer 2 6.068 3.03402 7.8628 0.0006999 \*\*\*  
## Residuals 93 35.886 0.38587   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

%95 güven düzeyinde söylenebilir ki gübreler arası farklılık vardır.

### Post-Hoc testi:

ANOVA=aov(rez\_aov)  
# Tukey test to study each pair of treatment :  
TukeyHSD(x=ANOVA, 'crop\_data$fertilizer', conf.level=0.95)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = rez\_aov)  
##   
## $`crop\_data$fertilizer`  
## diff lwr upr p adj  
## 2-1 0.1761687 -0.19371896 0.5460564 0.4954705  
## 3-1 0.5991256 0.22923789 0.9690133 0.0006125  
## 3-2 0.4229569 0.05306916 0.7928445 0.0208735

2 ile 1 ….. fark anlamlı değil  
3 ile 1 ….. fark anlamlı  
3 ile 2 ….. fark anlamlı

### Özetleyici istatistikler:

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

summary(crop\_data[crop\_data$fertilizer==1,])

## fertilizer yield   
## 1:32 Min. :175.4   
## 2: 0 1st Qu.:176.1   
## 3: 0 Median :176.8   
## Mean :176.8   
## 3rd Qu.:177.2   
## Max. :178.4

summary(crop\_data[crop\_data$fertilizer==2,])

## fertilizer yield   
## 1: 0 Min. :175.7   
## 2:32 1st Qu.:176.5   
## 3: 0 Median :177.0   
## Mean :176.9   
## 3rd Qu.:177.4   
## Max. :178.1

summary(crop\_data[crop\_data$fertilizer==3,])

## fertilizer yield   
## 1: 0 Min. :176.3   
## 2: 0 1st Qu.:177.0   
## 3:32 Median :177.3   
## Mean :177.4   
## 3rd Qu.:177.7   
## Max. :179.1

Post Hoc testlerinden yararlanılarak %95 güvenle söylenebilir ki 3 numaralı gübre en iyisidir.

## Tekrarlı Ölçümlü ANOVA:

Verileri içeri aktarma.

library(haven)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.5 v stringr 1.4.0  
## v tidyr 1.1.4 v forcats 0.5.1  
## v readr 2.0.2

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.1.2

library(rstatix)

## Warning: package 'rstatix' was built under R version 4.1.2

##   
## Attaching package: 'rstatix'

## The following object is masked from 'package:stats':  
##   
## filter

rept\_data <- read\_sav("~/GitHub/StatMeth2/Statistical-Methods-2/Repetitive\_ANOVA/repeated\_measures.sav")  
head(rept\_data,5)

## # A tibble: 5 x 4  
## TIME1 TIME2 TIME3 TIME4  
## <dbl> <dbl> <dbl> <dbl>  
## 1 65 66 63 63  
## 2 69 66 62 65  
## 3 60 64 59 59  
## 4 57 54 59 58  
## 5 62 62 62 54

### Verileri manipüle etme:

rept\_data[,5] <- list(seq(1,30,1))  
colnames(rept\_data) <- c("TIME1","TIME2","TIME3","TIME4","id")  
  
rp\_factor <- rept\_data %>%  
 gather(key = "time", value = "score", TIME1, TIME2, TIME3, TIME4) %>%  
 convert\_as\_factor(id,time)  
   
head(rp\_factor, 3)

## # A tibble: 3 x 3  
## id time score  
## <fct> <fct> <dbl>  
## 1 1 TIME1 65  
## 2 2 TIME1 69  
## 3 3 TIME1 60

### Ortalama ve St. Sapma değerleri:

rp\_factor%>%  
 group\_by(time) %>%  
 get\_summary\_stats(score, type = "mean\_sd")

## # A tibble: 4 x 5  
## time variable n mean sd  
## <fct> <chr> <dbl> <dbl> <dbl>  
## 1 TIME1 score 30 65.8 9.23  
## 2 TIME2 score 30 65.4 10.7   
## 3 TIME3 score 30 63.1 10.7   
## 4 TIME4 score 30 61.9 12.6

### Extreme Aykırı değer kontrolü:

rp\_factor %>%  
 group\_by(time) %>%  
 identify\_outliers(score)

## # A tibble: 11 x 5  
## time id score is.outlier is.extreme  
## <fct> <fct> <dbl> <lgl> <lgl>   
## 1 TIME2 11 85 TRUE FALSE   
## 2 TIME2 18 85 TRUE FALSE   
## 3 TIME2 28 41 TRUE FALSE   
## 4 TIME3 8 83 TRUE FALSE   
## 5 TIME3 11 83 TRUE FALSE   
## 6 TIME3 18 84 TRUE FALSE   
## 7 TIME3 23 78 TRUE FALSE   
## 8 TIME3 24 80 TRUE FALSE   
## 9 TIME3 25 42 TRUE FALSE   
## 10 TIME3 28 43 TRUE FALSE   
## 11 TIME4 24 91 TRUE FALSE

“EXTREME” bir aykırı değer yok.

### Normallik testi:

H0: Verilerin dağılımı ile normal dağılım arasında fark yoktur.  
Hs: Verilerin dağılımı ile normal dağılım arasında fark vardır.

rp\_factor %>%  
 group\_by(time) %>%  
 shapiro\_test(score)

## # A tibble: 4 x 4  
## time variable statistic p  
## <fct> <chr> <dbl> <dbl>  
## 1 TIME1 score 0.963 0.364   
## 2 TIME2 score 0.962 0.358   
## 3 TIME3 score 0.936 0.0697  
## 4 TIME4 score 0.966 0.432

Tüm değişkenler adına p>0.05 için H0 reddedilemez. Hepsi normal dağılıyor denebilir.

### Varyans homojenliği:

H0: 1 = 2 = 3 = 4  
Hs: En az bir i farklıdır.

bartlett.test(score~time,data=rp\_factor)

##   
## Bartlett test of homogeneity of variances  
##   
## data: score by time  
## Bartlett's K-squared = 2.7425, df = 3, p-value = 0.433

Test sonucuna göre p > 0.05 için söylenebilir ki varyanslar homojendir.

### ANOVA testi:

H0: Zamanlar arası fark yoktur.  
Hs: En az bir zamana ait değerler farklıdır.

rp\_factor$time <- as.factor(rp\_factor$time)  
rez\_aov=lm( rp\_factor$score ~ rp\_factor$time)   
anova(aov(rez\_aov))

## Analysis of Variance Table  
##   
## Response: rp\_factor$score  
## Df Sum Sq Mean Sq F value Pr(>F)  
## rp\_factor$time 3 310.7 103.58 0.8789 0.4543  
## Residuals 116 13670.7 117.85

[(Fh=0.878) < (Ft=3,15)] aynı zamanda Pvalue=0.4543 > alpha=0.05 için;  
H0 hipotezi reddedilemez. Yani zamanlar arasında farklılık yoktur.

## Rastgele Blok:

verileri içeri aktarma

library(readxl)  
cb\_data<- read\_excel("~/GitHub/StatMeth2/Statistical-Methods-2/Randomized\_Block\_ANOVA/castle\_bakery\_data.xlsx")  
colnames(cb\_data) <- c("Heights","Display\_regular","Display\_wide")  
head(cb\_data)

## # A tibble: 6 x 3  
## Heights Display\_regular Display\_wide  
## <chr> <dbl> <dbl>  
## 1 Bottom 47 46  
## 2 Bottom 43 40  
## 3 Middle 62 67  
## 4 Middle 68 71  
## 5 Top 41 42  
## 6 Top 39 46

### Normallik testi:

H0: Verilerin dağılımı ile normal dağılım arasında fark yoktur.  
Hs: Verileirn dağılımı ile normal dağılım arasında fark vardır.

shapiro.test(cb\_data$Display\_regular)

##   
## Shapiro-Wilk normality test  
##   
## data: cb\_data$Display\_regular  
## W = 0.85126, p-value = 0.1611

shapiro.test(cb\_data$Display\_wide)

##   
## Shapiro-Wilk normality test  
##   
## data: cb\_data$Display\_wide  
## W = 0.80521, p-value = 0.06544

%95 güven düzeyinde, bütün oynatma türü verileri normal dağılıyor denebilir.

### Verileri manipüle etme:

cb\_data <- cb\_data %>%  
 gather(key = "kind", value = "value", Display\_regular, Display\_wide) %>%  
 convert\_as\_factor(Heights,kind)

### Varyans homojenliği:

H0: Bottom = Middle = Top  
Hs: En az bir i farklıdır.

bartlett.test(value ~ Heights, data=cb\_data)

##   
## Bartlett test of homogeneity of variances  
##   
## data: value by Heights  
## Bartlett's K-squared = 0.16227, df = 2, p-value = 0.9221

P-value > 0.05 için H0 reddedilemez. Bu nedenle söylenebilir ki varyanslar homojendir.

### ANOVA testi:

H0: Plakların oynatma yerleri arasında fark yoktur.  
Hs: Plakların oynatma yerleri arasında fark vardır.

fit <- aov(value ~ Heights + kind, data = cb\_data)  
summary(fit)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Heights 2 1544 772.0 71.814 7.75e-06 \*\*\*  
## kind 1 12 12.0 1.116 0.322   
## Residuals 8 86 10.8   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

P < 0.05 için H0 reddedilir. Yani oynatma yerleri arasında farklılık vardır.

### Post Hoc testi:

TukeyHSD(fit ,conf.level=0.95)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = value ~ Heights + kind, data = cb\_data)  
##   
## $Heights  
## diff lwr upr p adj  
## Middle-Bottom 23 16.375288 29.624712 0.0000236  
## Top-Bottom -2 -8.624712 4.624712 0.6771042  
## Top-Middle -25 -31.624712 -18.375288 0.0000126  
##   
## $kind  
## diff lwr upr p adj  
## Display\_wide-Display\_regular 2 -2.365195 6.365195 0.3215794

Middle ile Bottom ….. fark anlamlı.  
Top ile Bottom ….. fark anlamlı değil.  
Top ile Middle ….. fark anlamlı  
Tukey testinin sonucuna göre söylenebilir ki “Middle” seviyesindeki oynatma en iyisidir.

## Latin Kare:

Verileri içeri aktarma:

library(readxl)  
data = read\_excel(path = "C:/Users/lenovo/Documents/GitHub/StatMeth2/Statistical-Methods-2/Latin\_Square/data\_ls.xlsx",  
 col\_names = TRUE)  
  
head(data,5)

## # A tibble: 5 x 4  
## Row Column Varieties Yield  
## <dbl> <dbl> <chr> <dbl>  
## 1 1 1 B 1.64  
## 2 1 2 D 1.21  
## 3 1 3 C 1.42  
## 4 1 4 A 1.34  
## 5 2 1 C 1.48

#Veri yapısına bakılıp satır sütun görselleştirmesi yapıldığında:   
#---1-2-3-4-   
#-1-B-D-C-A-   
#-2-C-A-D-B-   
#-3-A-C-B-D-   
#-4-D-B-A-C-   
#Şeklindeki latin kare veri setimiz ortaya çıkmaktadır.

### Veriyi manipüle etme:

data$Row <- as.factor(data$Row)  
data$Column <- as.factor(data$Column)  
data$Varieties <- as.factor(data$Varieties)  
str(data)

## tibble [16 x 4] (S3: tbl\_df/tbl/data.frame)  
## $ Row : Factor w/ 4 levels "1","2","3","4": 1 1 1 1 2 2 2 2 3 3 ...  
## $ Column : Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 4 1 2 ...  
## $ Varieties: Factor w/ 4 levels "A","B","C","D": 2 4 3 1 3 1 4 2 1 3 ...  
## $ Yield : num [1:16] 1.64 1.21 1.42 1.34 1.48 ...

### ANOVA testi:

attach(data)  
model <- lm(Yield ~ Row+Column+Varieties)  
anova(model)

## Analysis of Variance Table  
##   
## Response: Yield  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Row 3 0.03015 0.010052 0.4654 0.716972   
## Column 3 0.82734 0.275781 12.7692 0.005148 \*\*  
## Varieties 3 0.42684 0.142281 6.5879 0.025092 \*   
## Residuals 6 0.12958 0.021597   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

P değerlerine göre %95 güven düzeyinde;  
Satırlar arası farklılık yoktur.  
Sütunlar arası farklılık vardır.  
Çeşitler arası farklılık vardır yorumları yapılabilir.

### Ortalama karşılaştırma testi:

library(agricolae)

## Warning: package 'agricolae' was built under R version 4.1.2

# LSD test  
LSD.test(y = model,trt = "Varieties",DFerror = model$df.residual,MSerror = deviance(model)/model$df.residual,alpha = 0.05,group = TRUE,console = TRUE)

##   
## Study: model ~ "Varieties"  
##   
## LSD t Test for Yield   
##   
## Mean Square Error: 0.0215974   
##   
## Varieties, means and individual ( 95 %) CI  
##   
## Yield std r LCL UCL Min Max  
## A 1.46375 0.2386900 4 1.2839503 1.64355 1.185 1.670  
## B 1.47125 0.2095382 4 1.2914503 1.65105 1.290 1.665  
## C 1.06750 0.4426153 4 0.8877003 1.24730 0.660 1.475  
## D 1.33875 0.1795538 4 1.1589503 1.51855 1.180 1.565  
##   
## Alpha: 0.05 ; DF Error: 6  
## Critical Value of t: 2.446912   
##   
## least Significant Difference: 0.2542752   
##   
## Treatments with the same letter are not significantly different.  
##   
## Yield groups  
## B 1.47125 a  
## A 1.46375 a  
## D 1.33875 a  
## C 1.06750 b

# SNK test  
SNK.test(y = Yield,trt = Varieties,DFerror = model$df.residual,MSerror = deviance(model)/model$df.residual,alpha = 0.05,group = TRUE,console = TRUE)

##   
## Study: Yield ~ Varieties  
##   
## Student Newman Keuls Test  
## for Yield   
##   
## Mean Square Error: 0.0215974   
##   
## Varieties, means  
##   
## Yield std r Min Max  
## A 1.46375 0.2386900 4 1.185 1.670  
## B 1.47125 0.2095382 4 1.290 1.665  
## C 1.06750 0.4426153 4 0.660 1.475  
## D 1.33875 0.1795538 4 1.180 1.565  
##   
## Alpha: 0.05 ; DF Error: 6   
##   
## Critical Range  
## 2 3 4   
## 0.2542752 0.3188452 0.3597299   
##   
## Means with the same letter are not significantly different.  
##   
## Yield groups  
## B 1.47125 a  
## A 1.46375 a  
## D 1.33875 a  
## C 1.06750 b

SNK ve LSD test sonuçlarına göre söylenebilir A çeşidi en iyisidir.

## Çok Etkenli Deney Düzeni:

Verileri içeri aktarma.

library(readr)  
library(dplyr)  
lab\_data <- read\_table("~/GitHub/StatMeth2/Statistical-Methods-2/Multifactor\_Setup/JAHANMI2.DAT.txt",col\_types = cols(Lab = col\_factor(levels = c("1","2", "3", "4", "5", "6", "7", "8"))))  
  
head(lab\_data)

## # A tibble: 6 x 16  
## Id Lab Num Test Y X1 X2 X3 X4 Trt Set Llab Rep  
## <dbl> <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 1 1 1 609. -1 -1 -1 -1 1 1 0.75 1  
## 2 2 1 2 1 570. -1 -1 -1 -1 1 1 0.75 1  
## 3 3 1 3 1 690. -1 -1 -1 -1 1 1 0.75 1  
## 4 4 1 4 1 748. -1 -1 -1 -1 1 1 0.75 1  
## 5 5 1 5 1 618. -1 -1 -1 -1 1 1 0.75 1  
## 6 6 1 6 1 612. -1 -1 -1 -1 1 1 0.75 1  
## # ... with 3 more variables: Bat <dbl>, Sblab <dbl>, Set2 <dbl>

### Verileri manipüle etme:

Daha anlaşılabilir ve daha kolay işlenebilir hale getirmek için üzerinde inceleme yapacağımız değişkenleri ayırdık.

lab\_edited <- lab\_data %>%  
 summarize(Lab=Lab, X3=X3, Y=Y)  
head(lab\_edited,5)

## # A tibble: 5 x 3  
## Lab X3 Y  
## <fct> <dbl> <dbl>  
## 1 1 -1 609.  
## 2 1 -1 570.  
## 3 1 -1 690.  
## 4 1 -1 748.  
## 5 1 -1 618.

### ANOVA testi:

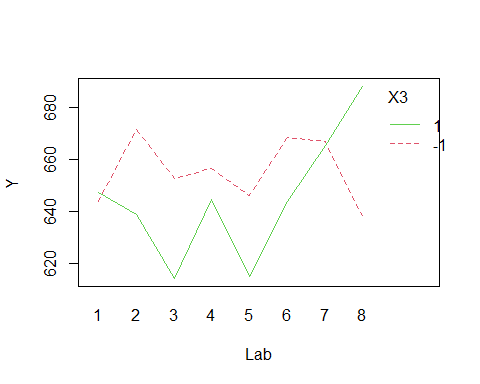
res\_1<- lm(Y ~ Lab\*as.factor(X3), data=lab\_edited)  
anova(res\_1)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Lab 7 70755 10108 1.8791 0.07113 .  
## as.factor(X3) 1 14380 14380 2.6733 0.10272   
## Lab:as.factor(X3) 7 87451 12493 2.3226 0.02453 \*  
## Residuals 464 2495861 5379   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

%95 Güven düzeyinde etkileşimler önemlidir.  
%95 Güven düzeyinde lab etkeni önemsizdir.  
%95 Güven düzeyinde X3 etkeni önemsizdir yorumları yapılabilir.

### Etkileşim grafikleri:

attach(lab\_edited)  
interaction.plot(Lab, X3, Y, type = "l",  
xlab = "Lab",trace.label = "X3",  
ylab = "Y", col=c(2,3,4))



## Kaynakça:

Tek Yönlü Anova Kaynağı: [Scribbr](https://www.scribbr.com)  
Tekrarlı Ölçümlü anova Kaynağı: [Google Drive](https://drive.google.com/file/d/0B0BsVnmXOaHwdzVYd05ZWjROQjg/view?resourcekey=0-yaqC7XcttKTgjEENapzhVg)  
Rastgele Blok Kaynağı: [National Institute of Standarts and Technology](https://www.nist.gov)  
Latin Kare Kaynağı: [Agron Info Tech](https://agroninfotech.blogspot.com)  
Çok Etkenli Deney Düzeni Kaynağı: [Information Technology Laboratory](https://www.nist.gov/itl)