Multivariate Statistical Analysis

Emircan ATALAY

Contents


```r
library(dplyr)
library(knitr)
library(magrittr)
library(knitr)
library(kableExtra)
library(car)
library(knitr)
library(rmarkdown)
```

SORULAR

SORU 1

Bir matrisin pozitif tanımlı olabilmesi için aşağıdaki şartları sağlaması gerekir.

1- Matris, kare matris olmalıdır.

2- Matris ile matrisin transpozu ($x = x^T$) birbirine eşit olmalıdır.

3- Matrisin tüm özdeğerleri pozitif olmalıdır.

A Matrisi için;

```r
A = matrix(c(5, -4, 3,
              4, 3, 2,
              3, 2, 2),
            nrow = 3, byrow = TRUE)
ta=t(A)
A
```

```
##      [,1] [,2] [,3]
## [1,]    5   -4    3
```

```
## [2,]    4    3    2
## [3,]    3    2    2
```

```r
ta
```

```
##      [,1] [,2] [,3]
## [1,]    5    4    3
## [2,]   -4    3    2
## [3,]    3    2    2
```

```r
combined_matrix = cbind(A, ta)

colnames(combined_matrix) = c("A1", "A2","A3", "ta1", "ta2","ta3")

kable(combined_matrix,
      align = c("l", "l", "l", "r", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "responsive")) %>%
  column_spec(1, border_left = TRUE) %>%
  column_spec(4, border_left = TRUE) %>%
  column_spec(ncol(combined_matrix), border_right = TRUE) %>%
  row_spec(0, bold = TRUE, background = "lightgray")
```

A1	**A2**	**A3**	**ta1**	**ta2**	**ta3**
5	-4	3	5	4	3
4	3	2	-4	3	2
3	2	2	3	2	2

A matrisi, transpozuna eşit olmadığı için simetrik bir matris değildir. Simetrik olmayan bir matris pozitif tanımlı bir matris olamaz. Bu sebeple A matrisi pozitif tanımlı bir matris **değildir**.

B Matrisi için;

```r
B = matrix(c(4, -3, 1,
             -3, 4, 2,
             1, 2, 3),
           nrow = 3, byrow = TRUE)

tb=t(B)

combined_matrix = cbind(B, tb)

colnames(combined_matrix) = c("B1", "B2","B3", "tb1", "tb2","tb3")

kable(combined_matrix,
      align = c("l", "l", "l", "r", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "responsive")) %>%
  column_spec(1, border_left = TRUE) %>%
  column_spec(4, border_left = TRUE) %>%
  column_spec(ncol(combined_matrix), border_right = TRUE) %>%
  row_spec(0, bold = TRUE, background = "lightgray")
```

B1	**B2**	**B3**	**tb1**	**tb2**	**tb3**
4	-3	1	4	-3	1

-3	4	2	-3	4	2
1	2	3	1	2	3

1.şart olan $B = B^T$ eşitliği sağlanıyor.

2.şart incelenirse;

```
result = eigen(B)
print(result$values)
```

```
## [1]  7.1464424  4.2184384 -0.3648808
```

B matrisinin özdeğerleri yukarıdaki gibi bulunmuştur. Bütün özdeğerler pozitif olmadığı için B matrisi pozitif tanımlı bir matris **değildir**.

C Matrisi için;

```
C = matrix(c(3, 2, 1,
              2, 6, 3,
              1, 3, 1),
            nrow = 3, byrow = TRUE)
tc=t(C)

combined_matrix = cbind(C, tc)

colnames(combined_matrix) = c("C1", "C2","C3", "tc1", "tc2","tc3")

kable(combined_matrix,
      align = c("l", "l", "l", "r", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "responsive")) %>%
  column_spec(1, border_left = TRUE) %>%
  column_spec(4, border_left = TRUE) %>%
  column_spec(ncol(combined_matrix), border_right = TRUE) %>%
  row_spec(0, bold = TRUE, background = "lightgray")
```

C1	C2	C3	tc1	tc2	tc3
3	2	1	3	2	1
2	6	3	2	6	3
1	3	1	1	3	1

1.şart olan $C = C^T$ eşitliği sağlanıyor.

2.şart incelenirse;

```
result = eigen(C)
print(result$values)
```

```
## [1]  8.3407707  2.0655400 -0.4063107
```

C matrisinin özdeğerleri yukarıdaki gibi bulunmuştur. Bütün özdeğerler pozitif olmadığı için C matrisi pozitif tanımlı bir matris **değildir**.

D Matrisi için;

```r
D = matrix(c(3, 3, 1,
              3, 4, 2,
              1, 2, 2),
            nrow = 3, byrow = TRUE)
td=t(D)

combined_matrix = cbind(D, td)

colnames(combined_matrix) = c("D1", "D2","D3", "td1", "td2","td3")

kable(combined_matrix,
      align = c("l", "l", "l", "r", "r", "r")) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "responsive")) %>%
  column_spec(1, border_left = TRUE) %>%
  column_spec(4, border_left = TRUE) %>%
  column_spec(ncol(combined_matrix), border_right = TRUE) %>%
  row_spec(0, bold = TRUE, background = "lightgray")
```

D1	**D2**	**D3**	**td1**	**td2**	**td3**
3	3	1	3	3	1
3	4	2	3	4	2
1	2	2	1	2	2

1.şart olan $D = D^T$ eşitliği sağlanıyor.

2.şart incelenirse;

```r
result = eigen(D)
print(result$values)
```

```
## [1] 7.4188327 1.3867702 0.1943972
```

D matrisinin özdeğerleri yukarıdaki gibi bulunmuştur. Bütün özdeğerler pozitif olduğu için D matrisi **pozitif tanımlı bir matristir**.

SORU 2

Verilen fonksiyon çok değişkenli normal dağılım fonksiyonudur.

Karesel form(Q) $= x_1^2 + 5x_2^2 + 2x_3^2 + 6x_2x_3 - 4x_1 - 8x_2 - 6x_3 + 9$ olarak verilmiş.

Karesel form üzerinden Σ matrisi aşağıdaki gibi yazılabilir.

```r
sigma = matrix(c(1, 0, 0,
                  0, 5, 3,
                  0, 3, 2),
            nrow = 3, byrow = TRUE)
sigma
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    5    3
## [3,]    0    3    2
```

Σ matrisinin tersi alınarak kovaryans matrisi aşağıdaki gibi bulunur.

```
covm= solve(sigma)
covm
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    2   -3
## [3,]    0   -3    5
```

ortalama vektörünü bulmak için Karesel form(Q)'da $x_1, x_2 ve x_3$ değişkenlerinin kısmi türevleri alınacak, ve üç denklem çözülerek ortalama vektörü elde edilecektir.

1. denklem => $(\partial Q)/x_1$ => $2x_1 - 4 = 0$
2. denklem => $(\partial Q)/x_2$ => $10x_2 + 6x_3 - 8 = 0$
3. denklem => $(\partial Q)/x_3$ => $6x_2 + 4x_3 - 6 = 0$

Birinci denklem sonucunda $x_1 = 2$ olarak elde edilmiştir.

İkinci denklem ve üçüncü denklem aşağıdaki şekilde çözülüp nihayi sonuç elde edilmiştir.

```
a <- matrix(c(10, 6,
              6,  4), nrow = 2, byrow = TRUE)
b <- c(8, 6)

sonuc <- solve(a, b)
x2 <- sonuc[1]
x3 <- sonuc[2]

mu <- c(2, x2, x3)
cat("Ortalama vektörü ( ):\n")
```

```
## Ortalama vektörü ( ):
```

```
print(mu)
```

```
## [1]  2 -1  3
```

En yüksek yayılıma sahip değişkene kovaryans matrisinin köşegenlerine bakarak karar verebiliriz. En büyük değere sahip köşegenin değeri 5'dir. Bu sebep ile en büyük yayılıma sahip değişkenin x_3 olduğu,

en düşük ortalamaya sahip değişkenin ise ortalama vektöründen yola çıkılarak x_2 (-1) olduğu söylenebilir.

SORU 3

Birden fazla bağımlı değişken olduğu için MANOVA testi ile bağımlı değişkenler açısından gübre türleri arasında fark olup olmadığı incelenecektir.

Varsayımlar:

1- Değişkenler normal dağılım gösterir.

2- Grup varyansları homojendir.

3- Değişkenler arasında yeterli düzeyde ilişki vardır.

Öncelikle veriyi aktaralım ve düzenleyelim:

```
library(readxl)
gubre = read_excel("C:/Users/emirc/Downloads/gubre.xlsx")
```

```
summary(gubre)
```

```
##      Grup            Bitki Boyu (cm) Yaprak Alanı (cm²) Biyokütle (g)
##  Length:45          Min.   :30.20   Min.   : 98.9      Min.   : 8.20
##  Class :character   1st Qu.:30.70   1st Qu.:100.8      1st Qu.: 8.50
##  Mode  :character   Median :45.40   Median :200.5      Median :15.80
##                     Mean   :42.06   Mean   :176.8      Mean   :14.96
##                     3rd Qu.:49.90   3rd Qu.:228.7      3rd Qu.:20.50
##                     Max.   :50.70   Max.   :232.1      Max.   :21.00
```

Daha rahat çalışabilmek adına excel dosyasında bulunan sütun isimlerini aşağıdaki şekilde güncelliyoruz.

```r
colnames(gubre) = c("grup","bitkiboyu","yaprakalani","kutle")
colnames(gubre)
```

```
## [1] "grup"        "bitkiboyu"   "yaprakalani" "kutle"
```

Değişkenlerin normal dağılıma uygunluk gösterip göstermediğini %95 güven düzeyinde tek tek inceleyelim:

H_{0i}: Veri ile normal dağılım arasında fark yoktur.

H_{1i}: Veri ile normal dağılım arasında fark vardır.

```r
shapirobitkiboy <- gubre %>%
  group_by(grup) %>%
  summarise(shapiro = shapiro.test(bitkiboyu)$p.value)

shapiroyaprak <- gubre %>%
  group_by(grup) %>%
  summarise(shapiro = shapiro.test(yaprakalani)$p.value)

shapirokutle <- gubre %>%
  group_by(grup) %>%
  summarise(shapiro = shapiro.test(kutle)$p.value)

shapirobitkiboy
```

```
## # A tibble: 3 x 2
##   grup     shapiro
##   <chr>      <dbl>
## 1 Kimyasal   0.576
## 2 Kontrol    0.710
## 3 Organik    0.136
```

```r
shapiroyaprak
```

```
## # A tibble: 3 x 2
##   grup     shapiro
##   <chr>      <dbl>
## 1 Kimyasal   0.659
## 2 Kontrol    0.538
## 3 Organik    0.198
```

```r
shapirokutle
```

```
## # A tibble: 3 x 2
##   grup     shapiro
##   <chr>      <dbl>
## 1 Kimyasal   0.428
## 2 Kontrol    0.261
## 3 Organik    0.734
```

Tüm p-value > 0.05 olduğu için verinin normal dağılım gösterdiği söylenebilir.

Değişkenlerin Varyans homojenliğini %95 güven düzeyinde inceleyelim:

```
gubre$grup <- factor(gubre$grup)

bartlettbitki <- bartlett.test(bitkiboyu ~ grup, data = gubre)
bartlettyaprak <- bartlett.test(yaprakalani ~ grup, data = gubre)
bartlettkutle <- bartlett.test(kutle ~ grup, data = gubre)

bartlettbitki
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  bitkiboyu by grup
## Bartlett's K-squared = 16.487, df = 2, p-value = 0.0002629
```

```
bartlettyaprak
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  yaprakalani by grup
## Bartlett's K-squared = 13.063, df = 2, p-value = 0.001456
```

```
bartlettkutle
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  kutle by grup
## Bartlett's K-squared = 4.2154, df = 2, p-value = 0.1215
```

İki durum için istatistiksel olarak varyans homojenliği sağlanmamaktadır. Bu durum MANOVA testi uygulandığı zaman nispeten güvensiz sonuçlar doğurabilir fakat tüm durumlar için analize devam edeceğiz.

MANOVA TESTİ

H_0: Gruplar arasında farklılık yoktur.

H_1: En az bir grup farklıdır.

```
manova <- manova(cbind(bitkiboyu, yaprakalani, kutle) ~ grup, data = gubre)
summary(manova)
```

```
##           Df Pillai approx F num Df den Df    Pr(>F)
## grup       2 1.9946   5002.4      6     82 < 2.2e-16 ***
## Residuals 42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value < 0.05 olduğu için en az bir grubun farklılık gösterdiğini söyleyebiliriz.

Farklılık gösteren grubu bulabilmek adına her bir bağımlı değişken için ANOVA testi uygulanacaktır.

1- Bitki Boyu

H_0: Gruplar arasında **bitki boyu** açısından anlamlı bir fark yoktur.

H_1: En az bir grup farklıdır.

```r
anovaboy = aov(gubre$bitkiboyu ~ gubre$grup)
summary(anovaboy)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## gubre$grup   2 3155.2  1577.6    9846 <2e-16 ***
## Residuals   42    6.7     0.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value $<$ 0.05 olduğundan dolayı H_0 **REDDEDİLİR**. Yani en az bir grubun (organik, kimyasal, kontrol) bitki boyu üzerinde istatistiksel olarak anlamlı bir etkisi vardır.

Hangi Grup veya grupların etkisi olduğunu Tukey testi ile inceleyelim.

```r
tukeyboy <- TukeyHSD(anovaboy)
tukeyboy
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = gubre$bitkiboyu ~ gubre$grup)
##
## $`gubre$grup`
##                       diff        lwr       upr p adj
## Kontrol-Kimyasal -19.660000 -20.015097 -19.30490     0
## Organik-Kimyasal  -4.766667  -5.121763  -4.41157     0
## Organik-Kontrol   14.893333  14.538237  15.24843     0
```

Her ikili grup için p-value <0.05 olduğundan dolayı üç grubun da bitki boyu üzerinde istatistiksel olarak anlamlı bir etki yaptığını söyleyebiliriz. Gruplar arasındaki farklara(diff) baktığımız zaman En büyük etkiyi yapan grubun **Kimyasal** olduğunu söyleyebİliriz.

Eğer bitki boyunun uzun olması isteniyorsa **Kimyasal** gübre kullanılması önerilir.

2- Yaprak Alanı

H_0: Gruplar arasında **yaprak alanı** açısından anlamlı bir fark yoktur.

H_1: En az bir grup farklıdır.

```r
anovayaprak <- aov(gubre$yaprakalani ~ gubre$grup)
summary(anovayaprak)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## gubre$grup   2 138430   69215   36221 <2e-16 ***
## Residuals   42     80       2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value $<$ 0.05 olduğundan dolayı H_0 **REDDEDİLİR**. Yani en az bir grubun (organik, kimyasal, kontrol) bitki boyu üzerinde istatistiksel olarak anlamlı bir etkisi vardır.

Hangi Grup veya grupların etkisi olduğunu Tukey testi ile inceleyelim.

```r
tukeyyaprak <- TukeyHSD(anovayaprak)
tukeyyaprak
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
```

```
## Fit: aov(formula = gubre$yaprakalani ~ gubre$grup)
##
## $`gubre$grup`
##                        diff        lwr        upr p adj
## Kontrol-Kimyasal -129.64000 -130.86633 -128.41367     0
## Organik-Kimyasal  -29.63333  -30.85966  -28.40701     0
## Organik-Kontrol   100.00667   98.78034  101.23299     0
```

Her ikili grup için p-value <0.05 olduğundan dolayı üç grubun da yaprak alanı üzerinde istatistiksel olarak anlamlı bir etki yaptığını söyleyebiliriz. Gruplar arasındaki farklara(diff) baktığımız zaman En büyük etkiyi yapan grubun **Kimyasal** olduğunu söyleyebİliriz.

Eğer yaprak alanının fazla olması isteniyorsa **Kimyasal** gübre kullanılması önerilir.

3- Biyokütle

H_0: Gruplar arasında **biyokütle** açısından anlamlı bir fark yoktur.

H_1: En az bir grup farklıdır.

```
anovakutle <- aov(gubre$kutle ~ gubre$grup)
summary(anovakutle)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## gubre$grup   2 1145.5   572.8   12778 <2e-16 ***
## Residuals   42    1.9     0.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value $<$ 0.05 olduğundan dolayı H_0 **REDDEDİLİR**. Yani en az bir grubun (organik, kimyasal, kontrol) biyokütle üzerinde istatistiksel olarak anlamlı bir etkisi vardır.

Hangi Grup veya grupların etkisi olduğunu Tukey testi ile inceleyelim.

```
tukeykutle <- TukeyHSD(anovakutle)
tukeykutle
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = gubre$kutle ~ gubre$grup)
##
## $`gubre$grup`
##                        diff        lwr        upr p adj
## Kontrol-Kimyasal -12.280000 -12.467822 -12.092178     0
## Organik-Kimyasal  -4.933333  -5.121156  -4.745511     0
## Organik-Kontrol    7.346667   7.158844   7.534489     0
```

Her ikili grup için p-value <0.05 olduğundan dolayı üç grubun da biyokütle üzerinde istatistiksel olarak anlamlı bir etki yaptığını söyleyebiliriz. Gruplar arasındaki farklara(diff) baktığımız zaman En büyük etkiyi yapan grubun **Kimyasal** olduğunu söyleyebİliriz.

Eğer biyokütlenin fazla olması isteniyorsa **Kimyasal** gübre kullanılması önerilir.

Genel Yorum

Verilen özellikler göz önünde bulundurulduğunda ve yapılan istatistiksel testler sonucunda, **kimyasal** gübre kullanımının verimliliği artırdığı söylenebilir.

SORU 4

Veri matrisinin aktarılması ve düzenlenmesi aşağıdaki şekilde yapılmıştır.

```
veri4 <- matrix(
  c(230, 181, 165, 150, 97, 192, 181, 189, 172, 170,
    125, 99, 97, 115, 120, 100, 80, 90, 95, 125,
    200, 55, 105, 85, 0, 150, 85, 120, 110, 130,
    109, 107, 98, 71, 82, 103, 111, 93, 86, 78
  ),
  nrow = 4, byrow = TRUE
)

rownames(veri4) <- c("Satış", "Fiyat", "Reklam", "Asistan")
colnames(veri4) <- paste0("Dönem", 1:10)
veri4
```

```
##         Dönem1 Dönem2 Dönem3 Dönem4 Dönem5 Dönem6 Dönem7 Dönem8 Dönem9 Dönem10
## Satış      230    181    165    150     97    192    181    189    172     170
## Fiyat      125     99     97    115    120    100     80     90     95     125
## Reklam     200     55    105     85      0    150     85    120    110     130
## Asistan    109    107     98     71     82    103    111     93     86      78
```

Hazırlanan veri matrisinin, kovaryans matrisini oluşturalım ve değişkenler arasındaki ilişkiyi yorumlayalım.

```
kovaryans <- cov(t(veri4))
print("Kovaryans Matrisi:")
```

```
## [1] "Kovaryans Matrisi:"
```

```
kovaryans
```

```
##               Satış      Fiyat    Reklam   Asistan
## Satış   1152.45556  -88.91111 1589.6667  301.6000
## Fiyat    -88.91111  244.26667  102.3333 -101.7556
## Reklam  1589.66667  102.33333 2915.5556  233.6667
## Asistan  301.60000 -101.75556  233.6667  197.0667
```

En büyük yayılıma sahip değişkenin 2915.5 değeri ile **gazetelerdeki reklam maliyeti** olduğu söylenebilir.

Gazetelerdeki reklam maliyeti ve satış arasında pozitif yönlü bir etki olduğunu da söyleyebiliriz. Yani ne kadar reklam verilirse satış miktarı da eş zamanlı olarak artıyor denilebilir.

Mağaza yöneticisinin iddia ettiği değişkenleri incelersek:

Fiyat değişimi ile kazak satışı arasında negatif korelasyon bulunuyor. Yani fiyat arttıkça satışın azaldığını söyleyebiliriz.

İki değişken arasındaki doğrusal ilişkiyi daha iyi görebilmek adına korelasyon matrisine bakalım.

```
korelasyon <- cor(t(veri4))
print("Korelasyon Matrisi:")
```

```
## [1] "Korelasyon Matrisi:"
```

```
korelasyon
```

```
##               Satış      Fiyat    Reklam    Asistan
## Satış     1.0000000 -0.1675760 0.8672280  0.6328673
## Fiyat    -0.1675760  1.0000000 0.1212619 -0.4637879
## Reklam    0.8672280  0.1212619 1.0000000  0.3082688
```

```
## Asistan  0.6328673 -0.4637879 0.3082688  1.0000000
```

Fiyat değişimi ile kazak satışı arasındaki korelasyon **-0.167** olarak bulunmuştur. Bu değer iki değişken arasında **negatif bir doğrusal ilişki** olduğunu gösterir fakat **güçlü bir ilişki değildir**. Yani bir değişken arttığında diğerinin bir miktar azalma eğiliminde olduğu söylenebilir.

Bu iki değişken arasındaki ilişkinin anlamlı olup olmadığını görebilmek için Pearson Korelasyon testini %95 güven düzeyi için uygulayalım.

```
cortest = cor.test(t(veri4)[, "Satış"], t(veri4)[, "Fiyat"])
print("Korelasyon Testi:")
```

```
## [1] "Korelasyon Testi:"
```

```
cortest
```

```
##
##  Pearson's product-moment correlation
##
## data:  t(veri4)[, "Satış"] and t(veri4)[, "Fiyat"]
## t = -0.48078, df = 8, p-value = 0.6435
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7211171  0.5165518
## sample estimates:
##       cor
## -0.167576
```

Test sonucuna baktığımızda p-value $>$ 0.05 olduğundan dolayı, korelasyonun istatistiksel olarak anlamlı olmadığı söylenebilir. Yani bu iki değişken arasındaki ilişki önemsizdir.Bu sebeple mağaza yöneticisinin iddiası geçerli değildir. KABUL EDİLEMEZ.

SORU 5

Çok değişkenli normal dağılıma uyduğu bilindiğine göre;

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}$$

f(x) yoğunluk fonksiyonu yukarıdaki şekilde olacaktır.

$x_1 = [x_1 \quad x_3]$

$f(x_1|x_2 = 25)$ bulunması isteniyor.

Verilen değerleri tanımlayalım (kovaryans matrisi dörde bölünüp aşağıdaki gibi tanımlanmıştır.);

```
mu5 = matrix(c(300, 30, 500
              ), nrow = 3, byrow = FALSE)
mu5t = t(mu5)

sigma11 = matrix(c(4000, 6000,
              80,  120), nrow = 2, byrow = TRUE)

sigma12 = matrix(c(80, 4
                   ), nrow = 2, byrow = FALSE)

sigma21 = matrix(c(6000, 10000
                   ), nrow = 1, byrow = FALSE)
```

```
sigma22 = matrix(c(120
                   ), nrow = 1, byrow = TRUE)

sigma11
```

```
##      [,1] [,2]
## [1,] 4000 6000
## [2,]   80  120
```

```
sigma12
```

```
##      [,1]
## [1,]   80
## [2,]    4
```

```
sigma21
```

```
##      [,1]  [,2]
## [1,] 6000 10000
```

```
sigma22
```

```
##      [,1]
## [1,]  120
```

$\mu = [300 \quad 500 \quad | \quad 25]$

$\theta = \mu_1' + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2)$

$$\theta = \begin{pmatrix} 300 \\ 500 \end{pmatrix} + \begin{pmatrix} 80 & 6000 \\ 4 & 120 \end{pmatrix} \ X \ (1/120) \ X \ (25 - 25) = \begin{pmatrix} 300 \\ 500 \end{pmatrix}$$

Koşullu ortalama vektörü yukarıdaki gibi bulunur.

Koşullu Kovaryans Matrisi:

$\Sigma_{kosul} = \Sigma_{[x_1,x_3]} - \Sigma_{[x_1,x_3,x_2]}\Sigma_{[x_2]}^{-1}\Sigma_{[x_2,x_1,x_3]}$

Formülü ile bulunur.

Hava sıcaklığının 25 derece olduğunun bilindiği bir gün için, dondurma sayısı ve günlük turist sayısı değişkenlerinin bileşik olasılık yoğunluk fonksiyonu ise aşağıdaki şekilde olacaktır.

$$f(x_1, x_3 \mid x_2 = 25) = \frac{1}{(2\pi)^{\frac{2}{2}} |\Sigma_{\text{cond}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 - 300 \\ x_3 - 500 \end{bmatrix}^T \Sigma_{\text{cond}}^{-1} \begin{bmatrix} x_1 - 300 \\ x_3 - 500 \end{bmatrix} \right)$$

SORU 6

Ders notlarının çok değişkenli normal dağıldığını varsayarsak:

$X \sim N(\mu, \Sigma)$

Y de normal dağılım gösterir diyebiliriz. Bu durumda normal dağılım özellikleri ile analiz edilebilir:

$Y \sim N(\mu_Y, \sigma_Y^2)$

```
notlar <- matrix(c(
  88, 76, 51,
  91, 82, 70,
  64, 72, 98,
```

```r
  83, 88, 66,
  77, 80, 94,
  45, 60, 55,
  48, 44, 73,
  51, 48, 64,
  63, 61, 79,
  65, 63, 44
), ncol = 3, byrow = TRUE)
```

Derslerin ortalama vektörü ve kovaryans matrisi aşağıdaki gibi hesaplanmıştır.

```r
mu6 = colMeans(notlar)
sigma6 = cov(notlar)
mu6
```

```
## [1] 67.5 67.4 69.4
```

```r
sigma6
```

```
##          [,1]      [,2]      [,3]
## [1,] 277.8333 213.88889   9.00000
## [2,] 213.8889 214.48889  44.37778
## [3,]   9.0000  44.37778 306.71111
```

Dersler arasında not ortalaması bakımından büyük fark olmadığı görülebilir. Fakat not ortalaması en yüksek olan ders **regresyon analizi**(69.4) dersidir diyebiliriz.

Kovaryans matrisine baktığımızda ise yayılımın en fazla olduğu dersin yine **regresyon analizi**(306.7) dersi olduğu görülmektedir

Eşit ağırlıklandırma söz konusu olduğu için ağırlıklar w isimli değişkene aşağıdaki gibi tanımlanmıştır. Sonrasında Y'nin Beklenen değer ve Varyansı aşağıdaki şekilde bulunmuştur.

```r
w <- c(1/3, 1/3, 1/3)
muY <- sum(w * mu6)
varyansY <- t(w) %*% sigma6 %*% w
ssY <- sqrt(varyansY)

muY
```

```
## [1] 68.1
```

```r
varyansY
```

```
##          [,1]
## [1,] 148.1741
```

```r
ssY
```

```
##          [,1]
## [1,] 12.17268
```

Y'nin 70'ten az olma olasılığı;

```r
deger = 70
z = (deger - muY) / ssY
olasilik <- pnorm(z)
cat("Y'nin 70'ten küçük olma olasılığı:",olasilik)
```

```
## Y'nin 70'ten küçük olma olasılığı: 0.5620179
```

%56 olarak bulunmuştur.