Multivariate HW1

Luke Beebe

2024-01-30

Using the iris data set that is available in R, find the mean vector Xbar(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) and the 4x4 Sample Variance-Covariance matrix.

First, we will load the data specified for this HW.

```
versicolor = iris[51:100,1:4]
head(versicolor)
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width
## 51          7.0         3.2          4.7         1.4
## 52          6.4         3.2          4.5         1.5
## 53          6.9         3.1          4.9         1.5
## 54          5.5         2.3          4.0         1.3
## 55          6.5         2.8          4.6         1.5
## 56          5.7         2.8          4.5         1.3
```

Now, we will calculate the mean vector Xbar using *apply()* which takes the data *versicolor*, specifies columns with *2*, and applies the *mean* function to each column of the dataset, returning Xbar.

```
apply(versicolor, 2, mean)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##        5.936        2.770        4.260        1.326
```

Next, we will calculate the 4x4 Sample Variance-Covariance matrix using *cov()*.

```
cov(versicolor)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length   0.26643265  0.08518367   0.18289796  0.05577959
## Sepal.Width    0.08518367  0.09846939   0.08265306  0.04120408
## Petal.Length   0.18289796  0.08265306   0.22081633  0.07310204
## Petal.Width    0.05577959  0.04120408   0.07310204  0.03910612
```

The Sample Variance-Covariance matrix finds the Covariance of each pair of variables by calculating (E[A * B] - E[A] * E[B]). Let's calculate the Covariance of *Sepal.Length* and *Sepal.Width* to check the matrix we produced above. It should equal **0.08518367**. One thing to keep in mind, is we are calculating the Sample Covariance and not the Population Covariance. We should divide by *n-1* and not *n*. You will see that in the equation below.

```
i=1
j=2
(mean(versicolor[,i]*versicolor[,j])-mean(versicolor[,i])*mean(versicolor[,j]))*length(versicolor[,i])/(length(ve
rsicolor[,i])-1)
```

```
## [1] 0.08518367
```

It checks out! Because I generalized the equation above, we could loop *i* and *j* to create our own Sample Variance-Covariance Matrix.