ST557: HOMEWORK 1

Brian Cervantes Alvarez
October 9, 2023

Question 1

```
# Set random seed for reproducibility
set.seed(503)

# Read HW1Q1 dataset in using base R
ds <- read.csv("HW1Q1.csv")
# Show first 5 rows
head(ds, 5)
```

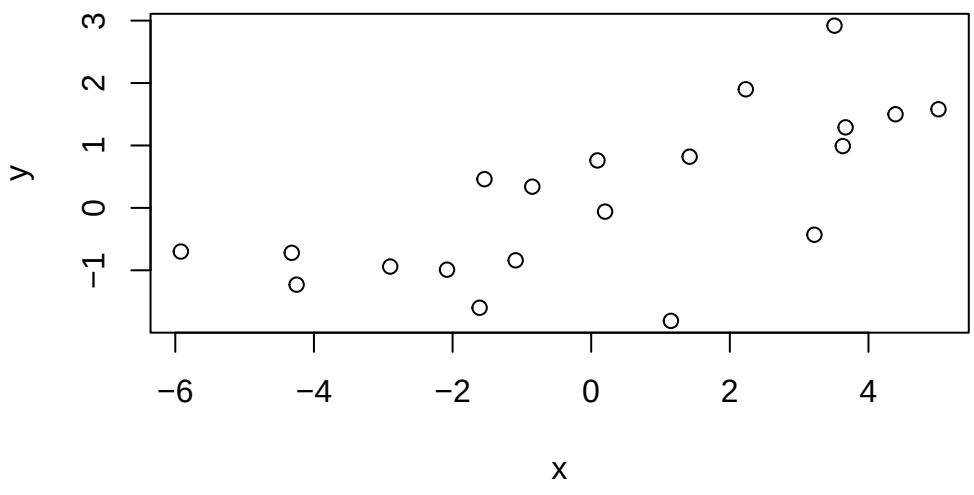
```
      X     Y
1 -1.54  0.46
2 -4.25 -1.23
3 -0.85  0.34
4 -2.90 -0.94
5 -1.09 -0.84
```

Part A

Read this data into R using read.csv(). Create a 2-dimensional scatter plot of the 20 observations (use plot() function in R).

```
# Get X and Y Components into vectors
x = ds$X
y = ds$Y

# Create scatter plot of Y ~ X
plot(x,y)
```



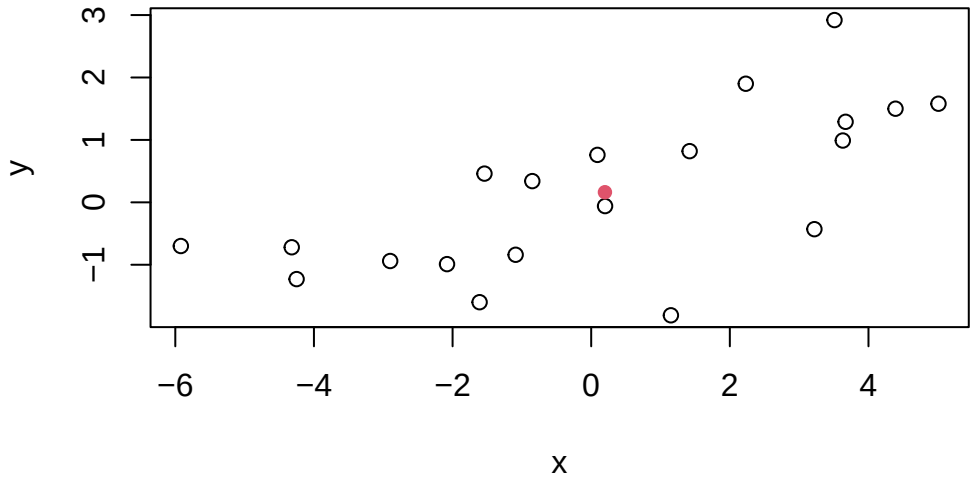
Part B

Find the sample mean vector, and add this point to the plot using points(). You can make this point a different color using the col= argument, or you can make it a different plotting character using the pch= argument. For example: > points(sampMean[1], sampMean[2], pch=16, col=2

```
plot(x,y)

sampMeanX = mean(x)
sampMeanY = mean(y)

points(sampMeanX, sampMeanY, pch=16, col=2)
```



Part C

Find the sample covariance matrix.

```
# Sample covariance matrix
sampCov <- cov(ds)

sampCov
```

```
          X        Y
X 10.140227 2.852078
Y  2.852078 1.668133
```

Part D

Find the eigendecomposition (spectral decomposition) of the sample covariance matrix using eigen().

```
sampEigenDecomp <- eigen(sampCov)

sampEigenDecomp
```

```
eigen() decomposition
$values
[1] 11.0108859  0.7974741

$vectors
           [,1]       [,2]
[1,] -0.9564274  0.2919702
[2,] -0.2919702 -0.9564274
```

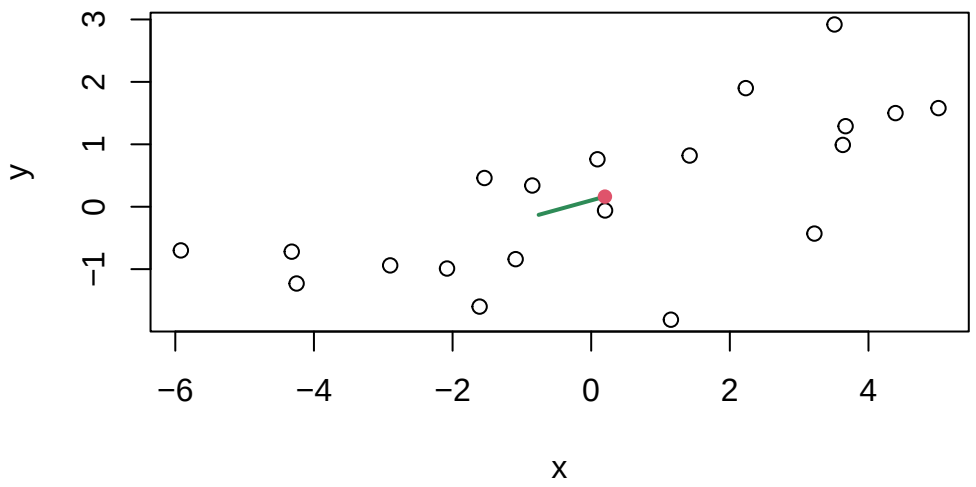
Part E

Add the eigenvector corresponding to the largest eigenvalue to the plot as a vector from the sample mean using `lines()`. Be careful here: if the first eigenvector is (v_1, v_2) and the sample mean vector is $(\bar{x}_1, \bar{x}_2)$, you want a line from $(\bar{x}_1, \bar{x}_2)$ to $(\bar{x}_1 + v_1, \bar{x}_2 + v_2)$."

```
# Find the index of the largest eigenvalue
largestEigenvalue <- which.max(sampEigenDecomp$values)
# Extract the eigenvector corresponding to the largest eigenvalue
largestEigenvector <- sampEigenDecomp$vectors[, largestEigenvalue]

# Calculate the endpoints of the line segment
sampMeanX <- mean(x)
sampMeanY <- mean(y)
xend <- sampMeanX + largestEigenvector[1]
yend <- sampMeanY + largestEigenvector[2]

plot(x,y)
# Add the line segment to the plot
lines(c(sampMeanX, xend), c(sampMeanY, yend), col = "seagreen4", lwd = 2)
points(sampMeanX, sampMeanY, pch=16, col=2)
```



Discussion for Part E

Question: Describe how the direction of this eigenvector relates to the cloud of data points

The eigenvector follows the trend of the data points, which themselves are showing a positive increasing trend, and it culminates precisely at the location of the sample means for X and Y.

Question 2