Matrix Algebra - Lab

Blake Pappas

Sep 15, 2022

Motor Trend Car Road Tests

The data was extracted from the 1974 Motor *Trend US* magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.

Load the Dataset

Code:

```
data(mtcars)
```

Question: What is the sample size of this data set and how many variables are there?

```
# Sample Size
(sample_size <- dim(mtcars)[1])
```

```
## [1] 32
```

```
# Number of Variables
(num_variables <- dim(mtcars)[2])
```

```
## [1] 11
```

Answer: The sample size of this data set is 32 and there are 11 variables.

Subset the Data

We will focus on the following variables:

1. mpg: Miles/gallon
2. disp: Displacement (cu.in.)
3. hp: Gross horsepower
4. drat: Rear axle ratio
5. wt: Weight (1000 lbs)

Question: Create a new data set from *mtcars* that only contains these 5 variables.

Code:

```r
vars <- which(names(mtcars) %in% c("mpg", "disp", "hp", "drat", "wt"))
(cars <- mtcars[, vars])
```

```
##                      mpg  disp  hp drat    wt
## Mazda RX4           21.0 160.0 110 3.90 2.620
## Mazda RX4 Wag       21.0 160.0 110 3.90 2.875
## Datsun 710          22.8 108.0  93 3.85 2.320
## Hornet 4 Drive      21.4 258.0 110 3.08 3.215
## Hornet Sportabout   18.7 360.0 175 3.15 3.440
## Valiant             18.1 225.0 105 2.76 3.460
## Duster 360          14.3 360.0 245 3.21 3.570
## Merc 240D           24.4 146.7  62 3.69 3.190
## Merc 230            22.8 140.8  95 3.92 3.150
## Merc 280            19.2 167.6 123 3.92 3.440
## Merc 280C           17.8 167.6 123 3.92 3.440
## Merc 450SE          16.4 275.8 180 3.07 4.070
## Merc 450SL          17.3 275.8 180 3.07 3.730
## Merc 450SLC         15.2 275.8 180 3.07 3.780
## Cadillac Fleetwood  10.4 472.0 205 2.93 5.250
## Lincoln Continental 10.4 460.0 215 3.00 5.424
## Chrysler Imperial   14.7 440.0 230 3.23 5.345
## Fiat 128            32.4  78.7  66 4.08 2.200
## Honda Civic         30.4  75.7  52 4.93 1.615
## Toyota Corolla      33.9  71.1  65 4.22 1.835
## Toyota Corona       21.5 120.1  97 3.70 2.465
## Dodge Challenger    15.5 318.0 150 2.76 3.520
## AMC Javelin         15.2 304.0 150 3.15 3.435
## Camaro Z28          13.3 350.0 245 3.73 3.840
## Pontiac Firebird    19.2 400.0 175 3.08 3.845
## Fiat X1-9           27.3  79.0  66 4.08 1.935
## Porsche 914-2       26.0 120.3  91 4.43 2.140
## Lotus Europa        30.4  95.1 113 3.77 1.513
## Ford Pantera L      15.8 351.0 264 4.22 3.170
## Ferrari Dino        19.7 145.0 175 3.62 2.770
## Maserati Bora       15.0 301.0 335 3.54 3.570
## Volvo 142E          21.4 121.0 109 4.11 2.780
```

Question: Compute the sample mean vector and the sample covariance matrix for this new data set.

Code:

```r
# Sample Mean Vector
n <- dim(cars)[1]
p <- dim(cars)[2]
X <- as.matrix(cars)
ones <- rep(1, n) # Creates a vector of 1's with a length of n
(meanCal <- (1 / n) * t(X) %*% ones)
```

```
##            [,1]
## mpg   20.090625
## disp 230.721875
## hp   146.687500
## drat   3.596563
## wt     3.217250
```

```r
# Sample Covariance Matrix
(S <- cov(cars))
```

```
##              mpg        disp         hp       drat          wt
## mpg    36.324103   -633.09721 -320.73206   2.1950635  -5.1166847
## disp -633.097208 15360.79983 6721.15867 -47.0640192 107.6842040
## hp   -320.732056  6721.15867 4700.86694 -16.4511089  44.1926613
## drat    2.195064   -47.06402  -16.45111   0.2858814  -0.3727207
## wt     -5.116685   107.68420   44.19266  -0.3727207   0.9573790
```

Plot the Data

Please summarize this new data set using only one graph that contains most of the information.

Code:

```r
# Chernoff Faces
library(aplpack)
par(mar = rep(0, 4))
faces(cars, cex = 0.8)
```


```
## effect of variables:
##  modified item       Var
##  "height of face   " "mpg"
```

```
## "width of face    " "disp"
## "structure of face" "hp"
## "height of mouth  " "drat"
## "width of mouth   " "wt"
## "smiling          " "mpg"
## "height of eyes   " "disp"
## "width of eyes    " "hp"
## "height of hair   " "drat"
## "width of hair   "  "wt"
## "style of hair   "  "mpg"
## "height of nose  "  "disp"
## "width of nose   "  "hp"
## "width of ear    "  "drat"
## "height of ear   "  "wt"
```

Multivariate Normal Distributions

In this section, we will simulate data from bivariate normal distributions with different correlation coefficients. Let's start with $\rho = 0$, that is, these two variables are independent to each other.

Code:

```
library(MASS)
n = 1000
rho <- 0
sigma <- 1
Sigma = matrix(c(sigma, sigma * rho,
                 sigma * rho, sigma), 2)
x1 <- mvrnorm(n = n, mu = c(0, 0), Sigma = Sigma)
```

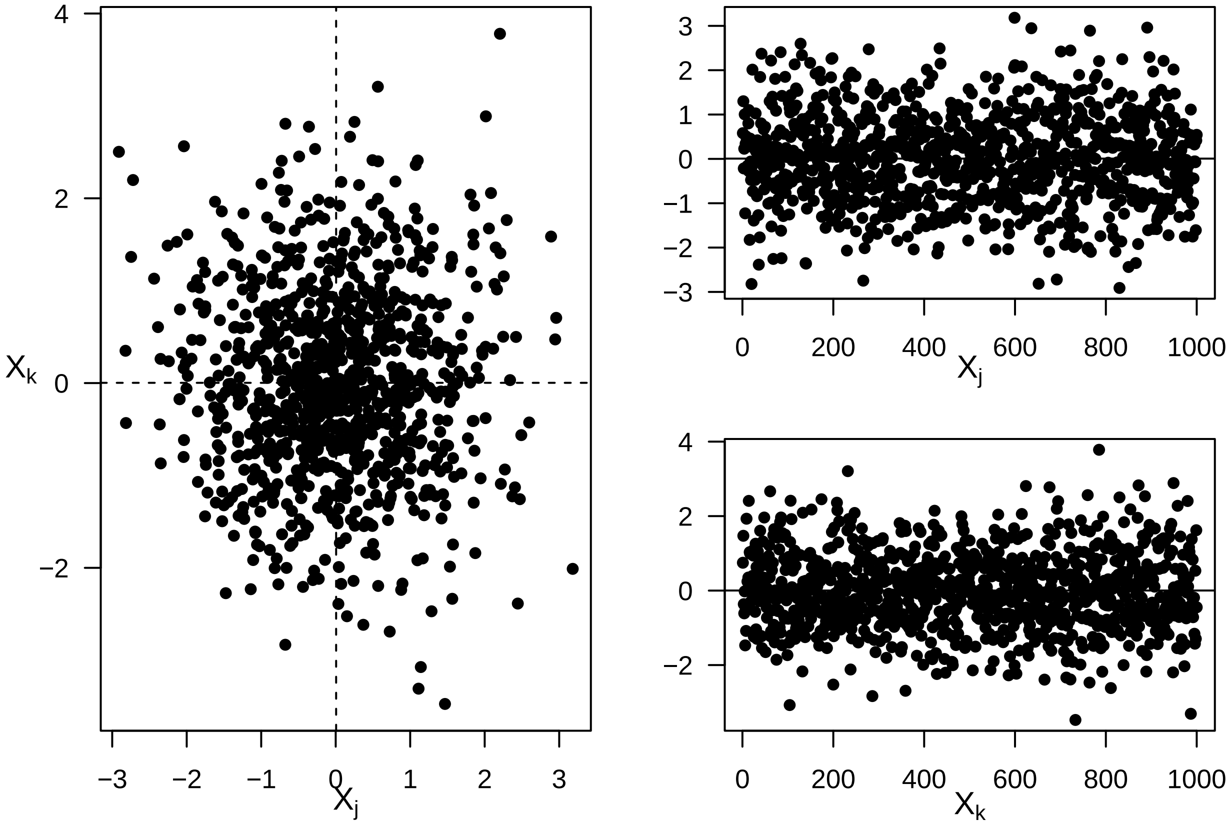
Plot This Simulated Data Set

Code:

```
par(mar = c(3.6, 3.6, 0.8, 0.6), las = 1)
layout(matrix(c(1, 1, 2, 3), nrow = 2, ncol = 2))
plot(x1, pch = 16, las = 1, xlab = "", ylab = "")
mtext(expression(X[j]), 1, line = 2); mtext(expression(X[k]), 2, line = 2)
text(-4, 2, expression(paste(S[jk], " = ")))
text(-3.3, 2, round(cov(x1[, 1], x1[, 2]), 2))
abline(h = mean(x1[, 2]), lty = 2); abline(v = mean(x1[, 1]), lty = 2)

plot(1:n, x1[, 1], pch = 16, xlab = "", ylab = "")
abline(h = mean(x1[, 1]))
mtext(expression(X[j]), 1, line = 2)

plot(1:n, x1[, 2], pch = 16, xlab = "", ylab = "")
abline(h = mean(x1[, 2]))
mtext(expression(X[k]), 1, line = 2)
```



Compute the Eigenvalues of the Sample Covariance Matrix

Code:

```
eigen <- eigen(S)

# Eigenvalues
(S %*% eigen$vectors[, 1] / eigen$vectors[, 1])
```

```
##          [,1]
## mpg  18636.79
## disp 18636.79
## hp   18636.79
## drat 18636.79
## wt   18636.79
```

```
# Eigenvectors
t(eigen$vectors[, 1]) %*% eigen$vectors[, 1]
```

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

```r
t(eigen$vectors[, 2]) %*% eigen$vectors[, 2]
```

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

```r
t(eigen$vectors[, 3]) %*% eigen$vectors[, 3]
```

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

```r
t(eigen$vectors[, 4]) %*% eigen$vectors[, 4]
```

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

```r
t(eigen$vectors[, 5]) %*% eigen$vectors[, 5]
```

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

Question: What are the theoretical eigenvalues and eigenvectors?

Answer: The theoretical eigennvectors are the direction in which the eigenvalues are stretched. The theoretical eigenvalues are the factor by which the eigenvectors are stretched.

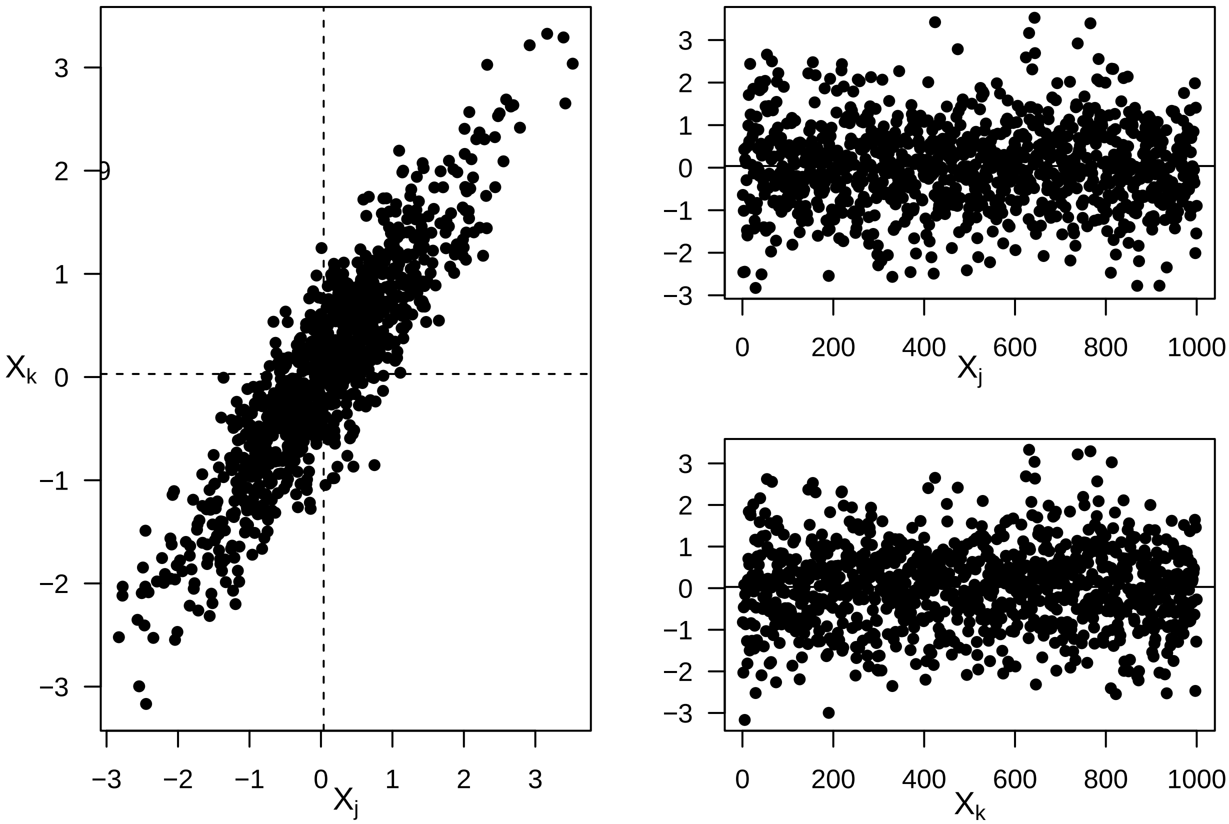
Repeat the Exercises Above but with $\rho = 0.9$

```r
library(MASS)
n = 1000
rho <- 0.9
sigma <- 1
Sigma = matrix(c(sigma, sigma * rho,
                 sigma * rho, sigma), 2)
x1 <- mvrnorm(n = n, mu = c(0, 0), Sigma = Sigma)
```

```r
par(mar = c(3.6, 3.6, 0.8, 0.6), las = 1)
layout(matrix(c(1, 1, 2, 3), nrow = 2, ncol = 2))
plot(x1, pch = 16, las = 1, xlab = "", ylab = "")
mtext(expression(X[j]), 1, line = 2); mtext(expression(X[k]), 2, line = 2)
text(-4, 2, expression(paste(S[jk], " = ")))
text(-3.3, 2, round(cov(x1[, 1], x1[, 2]), 2))
abline(h = mean(x1[, 2]), lty = 2); abline(v = mean(x1[, 1]), lty = 2)

plot(1:n, x1[, 1], pch = 16, xlab = "", ylab = "")
abline(h = mean(x1[, 1]))
mtext(expression(X[j]), 1, line = 2)

plot(1:n, x1[, 2], pch = 16, xlab = "", ylab = "")
abline(h = mean(x1[, 2]))
mtext(expression(X[k]), 1, line = 2)
```



3
2
1
0
−1
−2
−3
X_k
−3 −2 −1 0 1 2 3
X_j
3
2
1
0
−1
−2
−3
0 200 400 600 800 1000
X_j
3
2
1
0
−1
−2
−3
0 200 400 600 800 1000
X_k