Table of Contents

Problem 1

A

- How many rows are in a n×4 matrix?
- How many columns?
- Choose a number for n then give an example in R.

There are n rows and 4 columns in this matrix. If n = 2, here is a 2x4 matrix in R:

In [6]:
```
A1 <- matrix(c(3,2,5,6,2,7,5,8), ncol=4)
A1
```

A matrix: 2 × 4 of type dbl

3	5	2	5
2	6	7	8

B

- Give an example of a 4×2 matrix in R.

For a 4x2 matrix, we can take our previous example and tweak the number of columns:

In [7]:
```
B1 <- matrix(c(3,2,5,6,2,7,5,8), ncol=2)
B1
```

A matrix: 4 × 2 of type dbl

3	2
2	7
5	5
6	8

Problem 2

- A series of matrix multiplications are presented with the question of whether they are solvable
- Matrices are not printed because I'm struggling with the LaTeX package in Jupyter

A

This is solvable, the number of columns in the first matrix equals the number of rows in the second

B

This is also solvable, the number of columns is in the first matrix equals the number of rows in the second.

In R:

In [10]:
```
b2 <- matrix(c(1,2,-2,0,1,1), ncol=2)
b22 <- matrix(c(1,2,1,3,2,1), ncol=3)
Result3b = b2 %*% b22
Result3b
```

A matrix: 3 × 3 of type dbl

1	1	2
4	5	5
0	1	-3

C

This matrix multiplication is not solvable, there are not equal numbers of columns or rows in either equation.

In R:

In [11]:
```
C2 <- matrix(c(1,2,-2,0,1,1), ncol=2)
C22 <- matrix(c(1,-12,0), ncol = 1)
C2 %*% C22
```

```
Error in C2 %*% C22: non-conformable arguments
Traceback:
```

Problem 3

A

- Given matrix A, find the result of the multiplication: $(A)(A^-1)$. Is it solvable and why?

In [21]:
```
# Matrix A (A3)
A3 <- matrix(c(3,1,1,3), ncol=2)
A3
```

A matrix: 2 × 2 of type dbl

3	1
1	3

The matrix has an inverse, we can solve the equation and confirm by finding the determinant != 0:

In [17]:
```
A3inv = solve(A3)
round (A3 %*% A3inv, 10)
```

A matrix: 2 × 2 of type dbl

1	0
0	1

In [15]:
```
detA3 = det(A3)
detA3
```

8

B

- Given matrix A, find the result of the multiplication: $(A)(A^-1)$. Is it solvable and why?

In [18]:
```
# Matrix A (B3)
B3 <- matrix(c(3,1,0,0), ncol=2)
B3
```

A matrix: 2 × 2 of type dbl

3	0
1	0

This multiplication cannot be solved, we confirm by noting that the determinant = 0:

In [19]:
```
B3inv = solve(B3)
round (B3 %*% B3inv, 10)
```

```
Error in solve.default(B3): Lapack routine dgesv: system is exactly singular: U[2,2] = 0
Traceback:

1. solve(B3)
2. solve.default(B3)
```

In [20]:
```
det(B3)
```

0

Problem 4

This problem uses the USairpollution data from the HSAUR2 package. First we install and load the package, then assign the data to variable 'mydata'

In [22]:
```
install.packages("HSAUR2")
```

```
The downloaded binary packages are in
	/var/folders/1n/nvr79hb55tx9j41abrw6hsf80000gn/T//RtmpyP1gVX/downloaded_packages
```

In [23]:
```
library(HSAUR2)
```

```
Warning message:
"package 'HSAUR2' was built under R version 4.0.2"
Loading required package: tools
```

In [24]:
```
data("USairpollution", package = "HSAUR2")
mydata <- USairpollution
```

A

- Report the covariance and correlation matrix. Explain your findings.
- What variables are the most correlated, either positively or negatively? Why?

The most correlated variables (besides the correlation of the variables to themselves) are popul and manu with a correlation of .955 which is extremely high. The covariance measures are also very high, with the covariance of popul x manu being almost equal to the covariance of popul x popul.

In [27]:
```
cov(mydata)
```

A matrix: 7 × 7 of type dbl

	SO2	temp	manu	popul	wind	precip	predays
SO2	550.947561	-73.560671	8527.7201	6711.9945	3.1753049	15.0017988	229.92988
temp	-73.560671	52.239878	-773.9713	-262.3496	-3.6113537	32.8629884	-82.42616
manu	8527.720122	-773.971341	317502.8902	311718.8140	191.5481098	-215.0199024	1968.95976
popul	6711.994512	-262.349634	311718.8140	3353713.8939	175.9300610	-178.0528902	645.98598
wind	3.175305	-3.611354	191.5481	175.9301	2.0410244	-0.2185311	6.21439
precip	15.001799	32.862988	-215.0199	-178.0529	-0.2185311	138.5693840	154.79290
predays	229.929878	-82.426159	1968.9598	645.9860	6.2143902	154.7929024	702.59024

In [28]:
```
cor(mydata)
```

A matrix: 7 × 7 of type dbl

	SO2	temp	manu	popul	wind	precip	predays
SO2	1.00000000	-0.43360020	0.64476873	0.49377958	0.09469045	0.05429434	0.36956363
temp	-0.43360020	1.00000000	-0.19004216	-0.06267813	-0.34973963	0.38625342	-0.43024212
manu	0.64476873	-0.19004216	1.00000000	0.95526935	0.23794683	-0.03241688	0.13182930
popul	0.49377958	-0.06267813	0.95526935	1.00000000	0.21264375	-0.02611873	0.04208319
wind	0.09469045	-0.34973963	0.23794683	0.21264375	1.00000000	-0.01299438	0.16410559
precip	0.05429434	0.38625342	-0.03241688	-0.02611873	-0.01299438	1.00000000	0.49609671
predays	0.36956363	-0.43024212	0.13182930	0.04208319	0.16410559	0.49609671	1.00000000

B

- Report the covariance and correlation matrix among cities with the SO2 measure of more than 35.
- Explain your findings. What variables are the most correlated, either positively or negatively? Why?

Problem 5

- Convert this covariance matrix into the corresponding correlation matrix (using R).
- Also explain the intuition behind the difference between the covariance and correlation matrix.

In [31]:
```
Cov = matrix(c(3.8778, 2.8110, 3.1480, 3.5062,
               2.8110, 2.1210, 2.2669, 2.5690,
               3.1480, 2.2669, 2.6550, 2.8341,
               3.5062, 2.5690, 2.8341, 3.2352), byrow = T, nrow = 4)
```

The covariance matrix can be converted to the correlation matrix with `cov2cor()`. Covariance can be difficult to intuit because it depends on the scales or measures used by both variables. Correlation divides that covariance by the product of the variable standard deviations to give us a scale between -1 and 1 that makes the relationship easier to universally understand.

In [36]:
```
cov2cor(Cov)
```

A matrix: 4 × 4 of type dbl

1.0000000	0.9801619	0.9810921	0.9899048
0.9801619	1.0000000	0.9552780	0.9807159
0.9810921	0.9552780	1.0000000	0.9670131
0.9899048	0.9807159	0.9670131	1.0000000