1.

```
proc iml;
reset print;

A = {
     2 2,
    -1 1
};

B = {
     5  1,
     4 -2,
    -1  2
};

C = {
     6,
    -2,
     1
};

A_transposed = T(A);
B_transposed = T(B);
C_transposed = T(C);
```

1.a) C'B

```
Ctransposed_x_B = C_transposed * B;
```

21	12

1.b) A'B'

```
A_transposed_x_B_transposed = A_transposed * B_transposed;
```

9	10	-4
11	6	0

1.c) BA

```
B_x_A = B * A;
```

9	11
10	6
-4	0

1.d) BA

```
A_x_B = A * B;
```

- A row length = 2
- B col length = 3
  -> So it's not possible to multiply A * B

Also, sass gives us this error:

```
ERROR: (execution) Matrices do not conform to the operation.
```

QUESTION 2

a) Calculate G = F'F

```
F = {
  4 0 0,
  0 9 0,
  0 0 1
};

F_transpose = T(F);

G = F * F_transpose;
```

b) Eigenvalues

```
G_eigenvalues = eigval(G);
```


81,
16,
1

b) Eigenvectors

```
G_eigenvector = eigvec(G);
```

0	1	0
1	0	0
0	0	1

QUESTION 3.

You are given:

random vector

`X'= [ X1 , X2 , X3 , X4 ]` with

mean vector

μ_X = [4, 3, 2, 1]

variance-covariance matrix =

Σ_X =
[2 0 2 4,
2 1 9 2,
0 1 1 0,
3 0 2 2]

Partition X as:
x => X = [X_1, X_2, X_3, X_4] => [
X^a

...
X^b
]

X^a
=

Let A' = [1 –1] and Note: We have to `transpose` A to make it work

B = [2 -1, 0 1]

Consider the linear combinations $AX^{(1)}$ and BX^b to find:

a) $E(X^a) = \mu_x$

let X^a include X_1 and X_2

$\mu_1 = 4$
$\mu_2 = 3$

$E(X^a) = [4, 3]$ // *from mean vector*

b) $E(AX^a)$

$A^T = [1, -1] = [a, b]$

$\mu^1 = 4$
$\mu^2 = 3$

$E(AX^{(1)})$
$= a \times \mu^1 + b \times \mu^2$
= (1 x 4) + (3 x -1) // *by substitution*
= 1

c) $Cov(X^a)$

Cov => [3 0, 0 1]
the top left quartile of the covariance matrix grid

d) $Cov(AX^a)$

$Cov(A, X^a)$
$= ab\sigma_{12}$
= 1 x -1 x 0 *// by substitution*
= 0

e) $E(X^b)$

$E(X^b)$ = [2, 1] *// from mean vector*

f) $E(BX^b)$

// transpose:
B^T = [a, b] = { 2 0, -1 1 };

a' = [2 0]
b' = [-1 1]

$\mu_1 = 2$
$\mu_2 = 1$

$E(AX^a)$ = a x μ1 + b x μ2
// by substitution
= E('[2, 0] x 2) + ('[-1, 1] x 1)
= E[4, 0] + [-1, 1]
= E[3, 1]

g) $Cov(X^b)$

$Cov(X^b)$ = [
9 -2,
-2 4]
lower right quartile of covariance matrix

h) $Cov(BX^b)$

$Cov(B X^b) = B'Cov(X^b) B$

```
B = {
```

```
2 -1,
0  1
};

X_b = {
9 -2,
-2 4
};

H_Cov_B_Xb = T(B) * X_b * B;
```

36	-22
-22	17

i) Cov(X^a,X^b)

Formula:
Cov(X, Y) = E{(X - E(X))(Y - E(Y)^T)} = Cov(Y, X)^T

Cov(X^a, X^b) = E{(X^a - E(X^a))(X^b - E(X^b)^T)}

= E{(X^a - [4,3])(X^b - [2,1]^T)} *//substitution from (a) and (e)*

must be multiplied out to a 2 x 2 grid because it includes

[4,3] x [2 1]

And we are trying to find how X^a varies with X^b

Where the Variances intersect in the top right quartile of the variance covariance matrix

2	2
1	0

j) Cov(AX^a,BX^b)

$$\mathbf{Cov}(aX_1, bX_2) = \mathbf{E}[(aX_1 - a\mu_1)(bX_2 - b\mu_2)] = ab\mathbf{Cov}(X_1, X_2) = ab\sigma_{12}$$

From (i):

$Cov(X^a,X^b)$ =

2	2
1	0

a x b x $Cov(X^a,X^b)$

[1 –1] x [2 -1,0 1] x [2 2, 1 0]

```
qI_Cov_A_x_Xa_B_x_Xb = T(A) * B * {2 2, 1 0};
```

Output:

2	4

QUESTION 4

```
proc iml;

reset print;

data iceland;
infile "/folders/myfolders/sasuser.v94/iceland.csv" delimiter=",";
input TEMP PSAL DOXY NTRA PHOS SLCA;
run;
```

b) Produce the appropriate univariate descriptive statistics for each variable in the dataset using SAS code.

```
proc means data = iceland;
var TEMP PSAL DOXY NTRA PHOS SLCA;
run;
```

c) Choose an appropriate method to plot the dataset

```
proc sgplot data=iceland;
var TEMP PSAL DOXY NTRA PHOS SLCA;
scatter x=TEMP y=DOXY;
run;
```

d) Produce the covariance matrix for the dataset

```
proc CORR DATA=iceland COV;
var TEMP PSAL DOXY NTRA PHOS SLCA;
run;
```

e) Produce the correlation matrix for the dataset

```
proc corr data = iceland;
var TEMP PSAL DOXY NTRA PHOS SLCA;
run;
```

f) Using your answers from part b) to part e) above, summarise your exploration of the dataset and identify any potential issues arising from this exploration. (5 marks)

For the data iceland.csv which contains "information about ocean characteristics as collected by the International Council for the Exploration of the Sea (ICES)",

It has 145 records and the following variables:

- Temperature (Temp) [deg C]
- Salinity (PSal) [psu]
- Dissolved Oxygen (Doxy) [ml/l]
- Nitrate (Ntra) [umol/l]
- Phosphate (Phos) [umol/l]
- Silicate (SLCA) [umol/l]

The variables have the following means and standard deviations:

Data	MEAN	Standard Deviations

Data	MEAN	Standard Deviations
TEMP [deg C]	2.4573759	1.0831238
PSAL [psu]	34.3944586	0.3656668
DOXY [ml/l]	12.3627586	0.2582271
NTRA [umol/l]	0.9000000	0.4888968
PHOS [umol/l]	7.1359310	0.0371558
SLCA [umol/l]	8.5055172	1.7004811

- DOXY[ml/l] and TEMP[deg C] have a **negative** Pearson correlation of `-0.69838` with p `<.0001` and multiple outliers. See attached plot.

- The covariance matrix has 144 degrees of freedom.

- TEMP[deg C] highest **positive** covariance correlation is with SLCA[umol/l] of `0.565673060`

Potential issues arising from the dataset

- Covariance of Temperature with Silicate (SLCA) of -0.5656~ and covariance with Salinity -0.557~ are close, demonstrate that the Silicate and Salinity variables may not have bivariate independence.
- Unknown if data was not tested for normality, homoscedasticity
- Possible errors with measuring temperature and variables across locations

[iceland.csv]: ICES Dataset on Ocean Hydrography. The International Council for the Exploration of the Sea, Copenhagen. 2014