Bios 507 HW 1 due Feb 3

Problem 1

Let $Y \sim N(\overset{\text{mean}}{1.0}, \overset{\text{variance}}{2.0})$

1. What is $E(Y)$?

$E(Y) = \mu = 1.0$

2. What is $V(Y)$?

$V(Y) = 2.0$ because σ^2 is 2.0

3. What is $E(Y^2)$?

$V(Y) = E(Y^2) - [E(Y)]^2$
$-E(Y^2) \quad -E(Y^2)$

$V(Y) - E(Y^2) = -[E(Y)]^2$
$-V(Y) \qquad\qquad -V(Y)$

$-E(Y^2) = -V(Y) - [E(Y)]^2$
$\cdot(-1) \qquad \cdot(-1) \qquad \cdot(-1)$

$E(Y^2) = V(Y) + [E(Y)]^2$

$E(Y^2) = 2.0 + [1.0]^2$ plug in values now

$= E(Y^2) = 2.0 + 1.0$

$= E(Y^2) = 3.0$

Problem 2

Let x_1 and x_2 be 2 vectors defined as:

$$x_1 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} \quad x_2 = \begin{bmatrix} -1 \\ 2 \\ 9 \\ 3 \end{bmatrix}$$

checking if theres a relationship

1. Calculate the covariance between x_1 and x_2 using the summation formula.

$$\text{Cov}(x_1, x_2) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$$

step 1: compute the means ($\bar{x}_1$ & $\bar{x}_2$)

$$\bar{x}_1 = \frac{\overbrace{1+3}^{4} + \overbrace{5+7}^{12}}{4} = \frac{16}{4} = 4$$

$$\bar{x}_2 = \frac{\overbrace{-1+2}^{1} + \overbrace{9+3}^{14}}{4} = \frac{13}{4} = 3.25$$

step 2: center the data (aka: subtract mean)

$$x_1 = \begin{bmatrix} 1-4 = -3 \\ 3-4 = -1 \\ 5-4 = 1 \\ 7-4 = 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} -1-3.25 = -4.25 \\ 2-3.25 = -1.25 \\ 9-3.25 = 5.75 \\ 3-3.25 = -0.25 \end{bmatrix}$$

step 3: multiplying the centered values

$$(-3) \cdot (-4.25) = 12.75$$
$$(-1) \cdot (-1.25) = 1.25$$
$$(1) \cdot (5.75) = 5.75$$
$$(3) \cdot (-0.25) = -0.75$$

step 4: summing the products (this part Σ)

$$\Sigma = 12.75 + 1.25 + 5.75 - 0.75$$
$$= 19$$

step 5: dividing by $\frac{1}{n-1}$

$n = 4$ so $\frac{1}{3} \rightarrow \frac{19}{3} \Rightarrow 6.33$

$\text{Cov}(x_1, x_2) = 6.33$

Covariance is +ve so x_1 & x_2 will likely increase together!

Problem 2-2. Write the covariance between X_1 & X_2 using ① vector operations. Please write out the vector/matrix multiplication ② including the centering matrix, and then ③ show that the final answer is the same as what you obtained in the previous steps. (5)

Step 1: rewrite the vectors as matrices

→ bring back our centered vectors from 2-1

$$X_1 - \bar{X}_1 = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 3 \end{bmatrix} \quad X_2 - \bar{X}_2 = \begin{bmatrix} -4.25 \\ -1.25 \\ 5.75 \\ -0.25 \end{bmatrix}$$

→ Combine into just 1 matrix (centered still)

$$C = \begin{bmatrix} -3 & -4.25 \\ -1 & -1.25 \\ 1 & 5.75 \\ 3 & -0.25 \end{bmatrix}$$

→ each row is a single observation

→ each column is a different variable

Step 2: Matrix multiplication

a) first, transpose C, so C^T

$$C^T = \begin{bmatrix} -3 & -1 & 1 & 3 \\ -4.25 & -1.25 & 5.75 & -0.25 \end{bmatrix}$$

b) second do $C^T \cdot C$ to get the covariance matrix:

first column x first row

$C = [-3 \; -1 \; 1 \; 3] \times$
$C^T = [-3 \; -1 \; 1 \; 3]$

$(-3 \cdot -3) + (-1)(-1) + (1)(1) + (3)(3) = 20$

$(C^TC)_{1,1} = 20$

second column x first row

$C = [-4.25, -1.25, 5.75, -0.25]$

$C^T = [-3 \; -1 \; 1 \; 3]$

$((-4.25)(-3)) + ((-1.25)(-1)) + (5.75 \cdot 1) + ((-0.25) \cdot 3)$

$C^TC_{1,2} = 19$

$C^TC_{2,2} = \begin{matrix} [-4.25 & -1.25 & 5.75 & -0.25] \\ [-4.25 & -1.25 & 5.75 & -0.25] \end{matrix}$

$((-4.25) \cdot (-4.25)) + ((-1.25)(-1.25)) + (5.75 \cdot 5.75) + ((-0.25) \cdot (-0.25))$

$= 52.75$

$$C^TC = \begin{bmatrix} (1,1) & (1,2) \\ (2,1) & (2,2) \end{bmatrix}$$

symmetric!

$$C^TC = \begin{bmatrix} 20 & 19 \\ 19 & 52.75 \end{bmatrix}$$

this is our raw cov sum matrix

c) get the covariance value next page

(c) now get the covariance value

to calculate usi: $\boxed{\frac{1}{n-1} \cdot C^T C}$

don't do population!

$n=4$ so: $\frac{1}{3}$

this is our sample covariance matrix

$$\frac{1}{3}\begin{bmatrix} 20 & 19 \\ 19 & 52.75 \end{bmatrix}$$

$$\frac{20}{3} = 6.67 \quad \frac{19}{3} = 6.33, \quad \frac{52.75}{3} = 17.58$$

$$\begin{bmatrix} 6.67 & 6.33 \\ 6.33 & 17.58 \end{bmatrix}$$

So, $Cov(x_1, x_2) = 6.33$

which matches our 2-1 answer.

Problem 2-4

Using the covariance result you obtained earlier, evaluate between x_1 and x_2.

$$Corr(x_1, x_2) = \frac{Cov(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}}$$ (3)

→ first we need to get the standard deviation σ (2)

→ to get σ we need variance $V(x)$ first (1)

$$V(x_1) = \frac{1}{n-1} \sum (x_i - \bar{x}_1)^2 \Rightarrow \frac{1}{\cancel{4}3} \sum (x_i - 4)^2 \Rightarrow$$

$x_i - 4$	$(x_i - 4)^2$
$1-4=-3$	$(-3)^2 = 9$
$3-4=-1$	$(-1)^2 = 1$
$5-4=1$	$1^2 = 1$
$7-4=3$	$3^2 = 9$

$\Rightarrow 20$

$$\Rightarrow \frac{20}{4-1} \Rightarrow \frac{20}{3} = 6.67$$

$$V(x_2) = \frac{4}{3} \sum (x_i - \bar{x}_2)^2 \Rightarrow$$

$x_2 - 3.25$	$(x_2 - 3.25)^2$
-4.25	18.06
-1.25	1.56
5.75	33.06
-0.25	0.06

$$= \sum = 52.74 \Rightarrow \frac{52.74}{3} \Rightarrow 17.58$$

So, our $V(x_1) = 6.67$ & our $V(x_2) = 17.58$

Now, our $\sigma_{x_1} = \sqrt{6.67} = 2.58$ $\quad \sigma_{x_2} = \sqrt{17.58} = 4.19$

our sample covariance

Back to the formula: $Corr(x_1, x_2) = \frac{6.33}{2.58 \cdot 4.19}$ ← our σ_{x_1} & σ_{x_2}

$$= \frac{6.33}{10.81}$$

$\Rightarrow 0.585$ is the correlation between x_1 & x_2.

Problem 3

1. Write the assumed model:

$$y = \beta_0 + \beta_1 X + \epsilon$$

- $y \rightarrow$ returns for the second film $\rightarrow$ dependent variable
- $X \rightarrow$ returns for first film $\rightarrow$ Independent variable
- $\beta_0 \rightarrow$ Intercept
- $\beta_1 \rightarrow$ slope
- $\epsilon \rightarrow$ error term

2. Write the model matrix (it will be 4×2). rows columns

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} = \boxed{\begin{bmatrix} 1 & 90 \\ 1 & 85 \\ 1 & 87 \\ 1 & 103 \end{bmatrix}}$$

(ensuring we estimate β_0 & β_1)

Comes from 1 (β_0) (β_1)

3. In R.

4. How would you interpret β_1?

$\beta_1 = 1.1283$ so in our fitted model we have: $\hat{y} = -0.9606 + 1.1283X$, (predictions, β_0, β_1)
this means our slope is positive - there's likely a positive linear relationship here, where higher returns in the first film will likely garner larger returns on the second film. So, as X increases, y will increase.

5. How would you interpret β_0? Is this meaningful?

$\beta_0 = -0.9606$ so back to our fitted model, this intercept is negative, meaning at \$0 returned for a film the studio would be losing \$0.9606 million. Whether that's meaningful would depend on the industry technicalities, but a negative intercept would mean negative returns which you can't have sensically as the minimum is \$0.00. Therefore, this intercept isn't meaningful.

Problem 4

storage temperature (T) & # of weeks until food flavor deterioration starts (Y)

$$T = \begin{bmatrix} 8.0 \\ 4.0 \\ 0.0 \\ -4.0 \\ -8.0 \end{bmatrix} \qquad Y = \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix}$$

Assume a simple linear regression model (with intercept) is applicable.

1. Write the model matrix X.

$$X = \begin{bmatrix} 1 & 8.0 \\ 1 & 4.0 \\ 1 & 0.0 \\ 1 & -4.0 \\ 1 & -8.0 \end{bmatrix}$$

2. Find $Y'Y$ (Y transpose x Y).

$$Y' = [7.8 \;\; 9.0 \;\; 10.2 \;\; 11.0 \;\; 11.7] \qquad Y = \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix}$$

we are summing the squares!

$$Y'Y = (7.8 \times 7.8) + (9.0 \times 9.0) + (10.2 \times 10.2) + (11.0 \times 11.0) + (11.7 \times 11.7)$$

$$= 503.77$$

3. Evaluate $X'X$, where X is the model matrix. (X matrix transpose x X matrix)

first column ⇒ first row, then 2^{nd} column ⇒ 2^{nd} row to transpose

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8.0 & 4.0 & 0.0 & -4.0 & -8.0 \end{bmatrix} \cdot X = \begin{bmatrix} 1 & 8.0 \\ 1 & 4.0 \\ 1 & 0.0 \\ 1 & -4.0 \\ 1 & -8.0 \end{bmatrix}$$

1^{st} row & 1^{st} column
1^{st} row & 2^{nd} column
2^{nd} row & 2^{nd} column
2^{nd} row & 1^{st} column

kind of like FOIL

$$\Rightarrow (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) = 5$$

$$(1 \times 8) + (1 \times 4) + (1 \times 0) + (1 \times (-4)) + (1 \times (-8)) = 0$$

$$(8 \times 1) + (4 \times 1) + (0 \times 1) + (-4 \times 1) + (-8 \times 1) = 0$$

$$(8 \cdot 8) + (4 \cdot 4) + (0 \cdot 0) + (-4 \cdot -4) + (-8 \cdot -8) = 160$$

$$\Rightarrow \begin{bmatrix} 5 & 0 \\ 0 & 160 \end{bmatrix} = X'X$$

4. Evaluate $X'Y$.

We know X' & Y, now we need to multiply them

$$X' = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8.0 & 4.0 & 0.0 & -4.0 & -8.0 \end{bmatrix}}_{\text{2 x5 matrix}} \cdot Y = \underbrace{\begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix}}_{\text{5 x 1 matrix}} \Rightarrow$$

2 x 5 matrix
$\cdot$ $\Rightarrow$ 2 x 1 matrix
5 x 1 matrix

$$X'Y = (1 \cdot 7.8) + (1 \cdot 9.0) + (1 \cdot 10.2) + (1 \cdot 11.0) + (1 \cdot 11.7) = 49.7$$

$$(8.0 \cdot 7.8) + (4.0 \cdot 9.0) + (0 \cdot 10.2) + ((-4) \cdot 11.0) + ((-8) \cdot 11.7) = -39.2$$

$$\Rightarrow \boxed{\begin{bmatrix} 49.7 \\ -39.2 \end{bmatrix}}$$

5. Evaluate $(X'X)^{-1}$.

determinant

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$\begin{bmatrix} 5 & 0 \\ 0 & 160 \end{bmatrix}^{-1} = \frac{1}{(5 \cdot 160) - (\cancel{+}0) \cdot 160)} \begin{bmatrix} 160 & -(0) \\ -(0) & 5 \end{bmatrix}$$

$800 - 0$
800

$$= \frac{1}{800} \begin{bmatrix} 160 & 0 \\ 0 & 5 \end{bmatrix}$$

$$\boxed{\Rightarrow \begin{bmatrix} 0.2 & 0 \\ 0 & 0.00625 \end{bmatrix}}$$