
BIOS 507 HOMEWORK 1

Due 02/03/2025 by midnight

Directions: Complete all questions. Any R or SAS code used should be attached at the end of the homework. Collaboration is encouraged, but the final product must be your own work.

Problem 1

Let $Y \sim N(1.0, 2.0)$.

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

Problem 2

Let x_1 and x_2 be 2 vectors defined as:

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

1. Calculate the covariance between x_1 and x_2 using the summation formula.
2. Write the covariance between x_1 and x_2 using vector operations. Please write out the vector/matrix multiplication, including the centering matrix, and then show that the final answer is the same as what you obtained in the previous step.
3. Using either R or SAS PROC IML, create variables corresponding to x_1 and x_2 and evaluate their covariance.

- Using the covariance result that you obtained earlier, evaluate the correlation between x_1 and x_2 by hand.

Problem 3

A movie studio has started to produce many “two-part” films - releasing the second film a year or two after the first film. The studio has noticed that box office returns for the second movie are generally related to box office returns from the first movie and is interested in modelling this association so that they can predict future returns. The studio thinks that the box office returns for the second movie in each series can be modeled using a simple linear regression on the returns from the first movie. The studio collects the following data:

Movie Franchise	Returns For First Film (millions)	Returns For Second Film (millions)
A	90	120
B	85	90
C	87	87
D	103	111

- Write the assumed model
- Write the model matrix (it will be 4×2 .)
- Find the least squares solution for β_0 and β_1 using either SAS or R
- How would you interpret β_1 ?
- How would you interpret β_0 ? Is this meaningful?

Problem 4

The data shown below were obtained in a small scale experiment to study the relationship between storage temperature (T) and number of weeks before flavor deterioration of a food product begins to occur (Y).

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

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

- Write the model matrix, \mathbf{X}
- Find $\mathbf{Y}'\mathbf{Y}$ by hand
- Evaluate $\mathbf{X}'\mathbf{X}$ by hand, where \mathbf{X} is the model matrix.
- Evaluate $\mathbf{X}'\mathbf{Y}$ by hand.

5. Evaluate $(\mathbf{X}'\mathbf{X})^{-1}$ by hand. Note that this involves inverting a 2×2 matrix. You can find instructions on how to do this online (for example <https://www.mathsisfun.com/algebra/matrix-inverse.html>)