

Matrix Notation Practice

This practice is meant to get all students familiar with the very basic operations and terminology which we will use throughout the course. You can do this by hand or using a computer. Software for matrix algebra includes R, S-Plus, Matlab, and Mathematica.

1. *Matrix operations:* For the matrices below, find $A - B$, AC , and $B'A$.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 3 & 8 & 1 \\ 5 & 4 & 0 \end{bmatrix}$$

2. *Matrix characteristics: Linear dependence.*

- (a) Are the columns of A (below) linearly dependent? Justify your answer.

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 1 & 0 & 5 \\ 1 & 0 & 5 \end{bmatrix}$$

- (b) Is A an invertible matrix? Justify your answer.
 (c) Verify that C (below) is the inverse of B (below).

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ -5 & -3 & -4 \end{bmatrix} \quad C = \begin{bmatrix} -2 & -1 & -1 \\ 2 & -1 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

3. *Regression models in matrix notation:* An observational study is being done in youth smokers. Y_i is each youth's average number of cigarettes smoked per day. The first covariate being considered is the youth's age in years (so X_{i1} is the age). The second covariate being considered is whether or not the youth participates in after-school youth groups, such as athletic teams or community service groups (so X_{i2} is 1 for yes and 0 for no). Consider the following regression equations for the first 6 youth in the study ($i = 1, 2, 3, 4, 5, 6$):

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 * 13 && + \epsilon_1 \\ Y_2 &= \beta_0 + \beta_1 * 12 &+ \beta_2 &+ \epsilon_2 \\ Y_3 &= \beta_0 + \beta_1 * 16 && + \epsilon_3 \\ Y_4 &= \beta_0 + \beta_1 * 15 &+ \beta_2 &+ \epsilon_4 \\ Y_5 &= \beta_0 + \beta_1 * 12 && + \epsilon_5 \\ Y_6 &= \beta_0 + \beta_1 * 11 &+ \beta_2 &+ \epsilon_6 \end{aligned}$$

Re-write these equations as one matrix equation. Clearly define each matrix in your equation.

4. *Quadratic forms:* Suppose we determine that participation in youth groups is not a significant predictor of cigarette smoking, so the covariate X_{i2} is dropped from the model. Now we just have one covariate X_{i1} for age. Least squares (or maximum likelihood) estimation of a simple linear regression $Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$ is done by finding those $\hat{\beta}_0$ and $\hat{\beta}_1$ which minimize

$$\sum_{i=1}^m (Y_i - \beta_0 - \beta_1 X_{i1})^2$$

(equivalently, $\sum_{i=1}^m (Y_i - \hat{Y}_i)^2$). Re-write this summation in matrix notation.