Matrix Algebra for Statistics

What it's good for…

CLPS 2908 | Lecture 3 | January 31, 2019

1. Linear Dependence

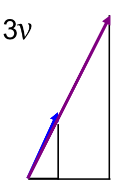
A matrix has true linear dependence when

- **pairs** of columns (or pairs of rows) are scalar multiplicatives of each other

  This means that column (row) vectors overlap (0°) and differ only in length.

$w = 3v$

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$



- or when one column (row) is a linear combination of **multiple other** columns (rows).

- True linear dependence is undesirable because it indicates **redundancy** (perfect correlations) in the data.
- The opposite of linear dependence is **orthogonality**, a perfect noncorrelation, or a 90°angle, between vectors.
- Most matrices have **correlations** among their columns (rows) that are *in between*: above 0° and below 90°.

Orthogonal: the *raw* vectors are at right angles;
Uncorrelated: the *centered* vectors are at right angles.

Rodgers, J. L., Nicewander, W. A., & Toothaker, L. (1984). Linearly Independent, Orthogonal, and Uncorrelated Variables. *The American Statistician, 38,* 134-135.

2. Rank

- Linear dependence ⇒ reduced **information density** in a matrix (because one row/column says nothing new).
- One way to assess this information density is with a matrix's ***rank***— the number of *independent* rows or columns (whichever is smaller).
  - 7 x 1 matrix has rank of 1
  - 2 x 3 matrix has rank of 2

Max rank = the smaller of the matrix's rows/columns

➢ Why important? Only **square** matrices of **full rank** (= "nonsingular" matrices) can be **inverted.** Inversion is a key operation in multivariate data analysis.

3. Inverses

- Is the matrix operation of **division.**

$$\frac{8}{2} = 8 \times \frac{1}{2} = 8 \times 2^{-1}$$

- What do you need inverses for?
  - To solve for unknowns, as in $\mathbf{A}x = \mathbf{b} \rightarrow x = \mathbf{A}^{-1}\mathbf{b}$
  - Later: to find "factor structures"
- How do you compute an inverse?
  - Via the determinant and "cofactors" (see separate handout)
  - With simultaneous equations

See Appendix 1

Computing an Inverse

a*a^{-1} = 1

$$\mathbf{A} \times \mathbf{A}^{-1} = \mathbf{I}$$

$$\begin{bmatrix} \boxed{1 \quad 3} \\ 4 \quad 5 \end{bmatrix} \times \begin{bmatrix} \boxed{\begin{matrix} w \\ y \end{matrix}} & \begin{matrix} x \\ z \end{matrix} \end{bmatrix} = \begin{bmatrix} \boxed{1} & 0 \\ 0 & 1 \end{bmatrix}$$

$$1w + 3y = 1$$

$$1w + 3y = 1$$
$$4w + 5y = 0$$
$$(4w + 12y = 4) - (4w + 5y = 0)$$
$$7y = 4, \text{ so } y = \frac{4}{7}$$
$$1w + 3(\frac{4}{7}) = 1$$
$$w = -\frac{5}{7}$$

Proceed similarly for the equations that include x and z, and you'll find that:

$$x = \frac{3}{7} \text{ and } z = -\frac{1}{7}. \text{ Thus,}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} -\frac{5}{7} & \frac{3}{7} \\ \frac{4}{7} & -\frac{1}{7} \end{bmatrix}$$

Note that all elements of the inverse are fractions with 7 as their denominator. 7 is his matrix's **determinant**. The numerators in the inverse elements, moreover, are a function of the diagonal elements of A, following an algorithm described in "determinant.pdf."

4. The Determinant

- More graded measure of information density: **determinant** ~ "generalized variance" of a matrix
  - picks up all the correlations among data vectors that are **not yet full** linear dependencies
  - is the "area" among data vectors See Appendix 2
  - **correlations** reduce that area, but **linear dependency** makes it go to 0
  - Limitation: can be computed only for **square matrices** *(e.g., var-cov)*, and Det > 0 only if matrix has **full rank**
  - indicates whether the matrix can be inverted:
  - If Det(**A**) = 0, inversion is not possible
    |**A**| = 0

Computing the Determinant

More details in handout

$$|\mathbf{A}| = \sum (-1)^{f(j_1, j_2, \ldots, j_p)} \prod_{i=1}^{p} a_{ij_i}$$

1. Permutating the columns of the matrix
2. Forming product of diagonal elements
3. Track number of changes that have been made = $f(j_1, j_2, \ldots, j_p)$
4. (-1) to the power of this number gives sign of product
5. Sum the signed products to find the determinant

$$\begin{bmatrix} 1 & 3 \\ 4 & 5 \end{bmatrix} \qquad \begin{bmatrix} 3 & 1 \\ 5 & 4 \end{bmatrix}$$

$$-1^0 \boxed{(1 \times 5)} + -1^1 \boxed{(3 \times 4)} = -7$$

$$\begin{bmatrix} 2 & -3 \\ 7 & 6 \end{bmatrix} \qquad \begin{bmatrix} -3 & 2 \\ 6 & 7 \end{bmatrix}$$

$$-1^0 (2 \times 6) + -1^1(-3 \times 7) = 33$$

5. Self-Multiplication

- Any n x p matrix **A** that is pre-multiplied by its own p x n transpose, **A′**, results in a p x p **sums-of-squares/cross-products** matrix of the p variables summed over the n cases:

A′ x **A** = **SSCP** of p vectors/variables

p x n n x p p x p

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -3 & -2 & -1 \end{bmatrix} \bullet \begin{bmatrix} 1 & 4 & -3 \\ 2 & 5 & -2 \\ 3 & 6 & -1 \end{bmatrix}$$

$1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 =$ **sum of squares**

$4 \cdot (-3) + 5 \cdot (-2) + 6 \cdot (-1) =$ **cross-product**

6. Derive S_{xx} and R_{xx} from X

We begin with the data matrix **X**: $\begin{bmatrix} 1 & 7 \\ 3 & 1 \\ 5 & 4 \end{bmatrix}$

First we want to turn **X** into "**Y**," the data matrix of mean-deviated (= centered) scores.

To that end we need a matrix of means.

Compute the sum by pre-multiplying by a unit vector: 1' **X**

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 7 \\ 3 & 1 \\ 5 & 4 \end{bmatrix} = \begin{bmatrix} 9 & 12 \end{bmatrix}$$

Now multiply by the scalar n^{-1}: $\boxed{\begin{bmatrix} 9 & 12 \end{bmatrix} n^{-1} = \begin{bmatrix} 3 & 4 \end{bmatrix}}$

The problem with 1' **X** n^{-1} is that its dimensionality is 1 x p.
But to create mean-deviated scores we need an n x p matrix.

We must expand by an appropriate unit vector that turns an 1 x p matrix into an n x p matrix. **How?**

Pre-multiply by an n x 1 unit vector **1**:

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 3 & 4 \end{bmatrix} = \boxed{\begin{bmatrix} 3 & 4 \\ 3 & 4 \\ 3 & 4 \end{bmatrix}}$$

1 (3x1)   **1' X**n^{-1} (1x2)   **11' X**n^{-1} (3x2)

Now we can subtract the matrix of means from the raw data matrix to yield **Y,** the matrix of mean-deviated scores

$$\begin{bmatrix} 1 & 7 \\ 3 & 1 \\ 5 & 4 \end{bmatrix} - \boxed{\begin{bmatrix} 3 & 4 \\ 3 & 4 \\ 3 & 4 \end{bmatrix}} = \begin{bmatrix} -2 & 3 \\ 0 & -3 \\ 2 & 0 \end{bmatrix}$$

Now, self-multiplication of the **Y** matrix yields a **SSQ/CP** matrix, and dividing by n-1 yields the **variance-covariance matrix**, $\mathbf{S}_{yy}$

$$\begin{bmatrix} -2 & 0 & 2 \\ 3 & -3 & 0 \end{bmatrix} \times \begin{bmatrix} -2 & 3 \\ 0 & -3 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix}$$

Y' (2x3) **Y** (3x2) **SS/CP** (2x2)

$$\begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix} \cdot (n-1)^{-1} = \mathbf{S}_{yy}$$

covariance

variances

How do we get from variances and covariances to **R**? We standardize. Correlations are covariances divided by the standard deviations of each variable:

$$\frac{cov_{x_1 x_2}}{s_1 s_2} = r_{x_1 x_2}$$

The trick is to find a diagonal matrix that **rescales** the entries in $\mathbf{S}_{yy}$ so that they are properly standardized.

This does the trick:

$$\mathbf{D}'_{s^{-1}} \cdot \mathbf{S}_{yy} \cdot \mathbf{D}_{s^{-1}} = \mathbf{R}_{xx}$$

S has y as subscripts because it is indeed the var-cov matrix of the mean-deviated Y scores. **R** has x as subscripts because it does show the correlation of the original X scores (as well as Y scores; for correlations that's equivalent)

$$\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix} \cdot \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix} \cdot \begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix} (n-1)^{-1} = \mathbf{R}_{xx}$$

The p**R**emultiplication by **D'** divides the entries in the first **R**ow by s_{x1} and the entries in the second **R**ow by s_{x2}.

The p**O**stmultiplication by **D** divides the entries in the first c**O**lumn by s_{x1} and the entries in the second column by s_{x2}.

Rescaling SSQ/CP Matrix

$$\boxed{\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix}} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{s_{x1}} & \longrightarrow \\ \longrightarrow & \frac{1}{s_{x2}} \end{bmatrix} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix}$$

Note: For graphical simplicity, I am leaving out the scalar $(n\text{-}1)^{-1}$, which should follow the SSQ/CP matrix.

$$\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{s_{x1}} & \longrightarrow \\ \longrightarrow & \frac{1}{s_{x2}} \end{bmatrix} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix}$$

$$\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{s_{x1}} \begin{bmatrix} 8 \\ -6 \end{bmatrix} \frac{1}{s_{x2}} \end{bmatrix} \begin{matrix} -6 \\ 18 \end{matrix}$$

$$\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix} \begin{bmatrix} 8 & -6 \\ -6 & 18 \end{bmatrix}$$

$$\begin{bmatrix} \frac{8}{\boxed{s_{x1}}} & \frac{-6}{\boxed{s_{x1}}} \\ \frac{-6}{\boxed{s_{x2}}} & \frac{18}{\boxed{s_{x2}}} \end{bmatrix}$$

$$\begin{bmatrix} \dfrac{8}{s_{x1}} & \dfrac{-6}{s_{x1}} \\ \dfrac{-6}{s_{x2}} & \dfrac{18}{s_{x2}} \end{bmatrix} \boxed{\begin{bmatrix} s_{x1}^{-1} & 0 \\ 0 & s_{x2}^{-1} \end{bmatrix}}$$

$$\begin{bmatrix} \dfrac{1}{s_{x1}} & \\ & \dfrac{1}{s_{x2}} \end{bmatrix}$$

$$\begin{bmatrix} \dfrac{8}{s_{x1}} & \dfrac{-6}{s_{x1}} \\ \dfrac{-6}{s_{x2}} & \dfrac{18}{s_{x2}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{s_{x1}} & \\ & \frac{1}{s_{x2}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{8}{s_{x1}} & \frac{-6}{s_{x1}} \\ \frac{-6}{s_{x2}} & \frac{18}{s_{x2}} \end{bmatrix}$$

$$\begin{bmatrix} \frac{8}{s_{x1}s_{x1}} & \frac{-6}{s_{x1}s_{x2}} \\ \frac{-6}{s_{x1}s_{x2}} & \frac{18}{s_{x2}s_{x2}} \end{bmatrix}(n-1)^{-1} = \mathbf{R}_{xx}$$

An alternative route to R: Standardizing the data in the first place

What is the formula for r, the correlation coefficient?

$$\frac{\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{s_1 s_2(n-1)}$$

Break it down, you get z values

$$\frac{\sum(X_1-\bar{X}_1)(X_2-\bar{X}_2)}{s_1 s_2(n-1)}$$

$$\sum z_i z_j \cdot (n-1)^{-1}$$

z-values are mean-deviated scores divided by their SD.

Thus, the correlation coefficient (formula above) is just the cross-multiplication of *z* values for two *different* variables.

$$\frac{\sum(X_1-\bar{X}_1)(X_1-\bar{X}_1)}{s_1 s_1(n-1)}$$

$\rightarrow s^2$

When we cross-multiply the *same* variable (X_1), we get $s^2/s_1s_1 = 1$.

Two Paths to R

Just now, we took this step:

$$\mathbf{Z} = \mathbf{YD}_{s^{-1}}$$

…and then this step:

$$\mathbf{R}_{xx} = \mathbf{Z'Z}(n-1)^{-1}$$

Substituting the left equation in the right equation, we get the equation we had before:

$$\mathbf{R}_{xx} = (\mathbf{YD}_{s^{-1}})'(\mathbf{YD}_{s^{-1}}) = \mathbf{D'}_{s^{-1}}\ \mathbf{Y'YD}_{s^{-1}} = \mathbf{D'}_{s^{-1}}\ \mathbf{S}_{yy}\ \mathbf{D}_{s^{-1}}$$

because $\mathbf{S}_{yy} = \mathbf{Y'Y}$

So you can (1) go from **X** via **Y'Y** to **S**, standardize, and get **R** (the earlier path); or (2) you can standardize first to yield **Z** and get **R** via **Z'Z** (the recent path).

The variance-covariance matrix **S** (or, standardized, the correlation matrix **R**) represent the core of most multivariate analyses (e.g., PCA, Manova, MRegression)

Appendix 1: Linear Combinations

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & 5 \end{bmatrix}$$

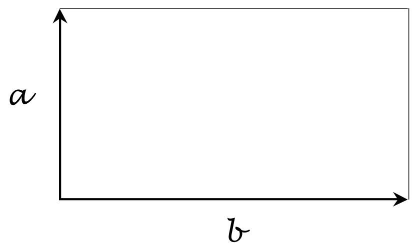
$$(2 * a_{i1}) - a_{i2} = a_{i3}$$

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & 1 \\ 3 & 1 & 5 \end{bmatrix}$$

$$(2 * a_{i1}) - a_{i2} = a_{i3}$$

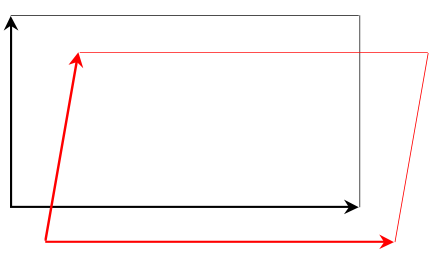
$$(5 * a_{2j}) - (7 * a_{1j}) = a_{i3}$$

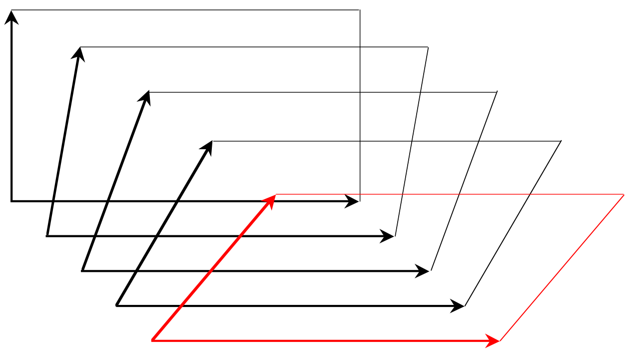
Appendix 2: Determinant as Area

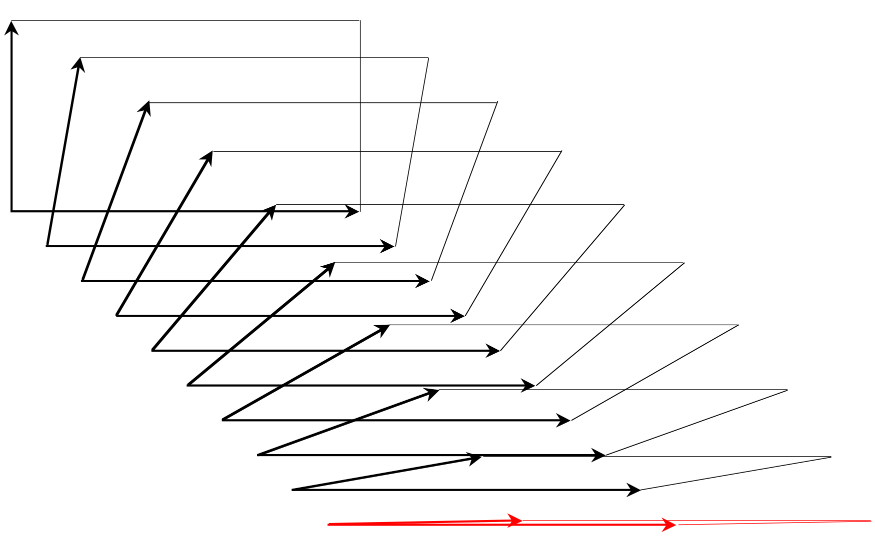


Determinant = area spanned
by the two vectors, a and b

Change angle between
a and b $\longrightarrow$ correlation







Determinant → 0, when the vector angle becomes 0°