• Approach 2: columns of $X$ as $p$ vectors in $n$-dimensional space (c.f. approach 1)

$$X_{(n \times p)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = [y_1 \mid y_2 \mid \cdots \mid y_p]$$

- benefit of approach 2: many of the algebraic expressions we shall encounter in multivariate analysis can be related to the geometrical notion of length, angle, volume.

- sample mean

$$y_i = \left( \frac{1}{\sqrt{n}} \right) \frac{1}{\sqrt{n}} = \frac{x_{1i} + x_{2i} + \cdots + x_{ni}}{n} = \bar{x}_1$$

- sample variance

$$d_i = y_i - \bar{x}_1 = \begin{bmatrix} x_{1i} - \bar{x}_1 \\ x_{2i} - \bar{x}_1 \\ \vdots \\ x_{ni} - \bar{x}_1 \end{bmatrix}$$

$$L_d^2 = d_i^T d_i = \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2$$

- sample covariance and correlation

$$d_i^T d_k = L_d L_{d_k} \cos(\theta_{ik}) = \sum_{j=1}^{n} (x_{ji} - \bar{x}_i)(x_{jk} - \bar{x}_k)$$

$$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii} s_{kk}}} = \cos(\theta_{ik})$$

- visualization of objects in 3-dim is useful to illustrate certain statistical concepts in terms of only 2 or 3 vectors of any $n$-dim

Reading: Textbook, 3.2, 3.3, 1.6