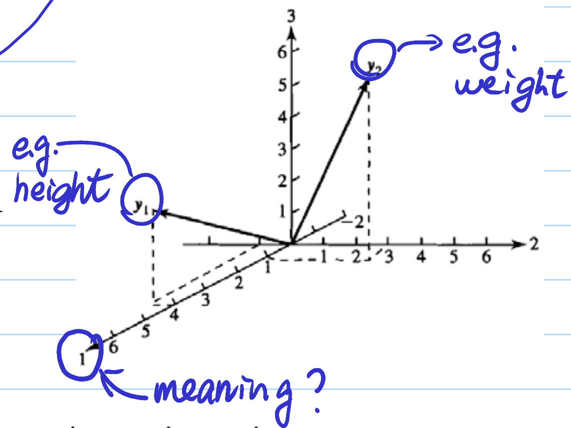
- Approach 2: columns of X as p vectors in n-dimensional space (c.f. approach 1)

Recall: Linear Model

$$\underset{(n\times p)}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = [\mathbf{y}_1 \mid \mathbf{y}_2 \mid \cdots \mid \mathbf{y}_p]$$

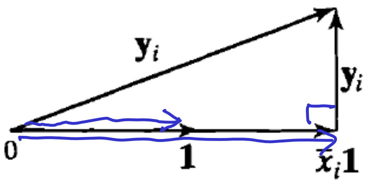


➢benefit of approach 2: many of the algebraic expressions we shall encounter in multivariate analysis can be related to the geometrical notion of *length*, *angle*, *volume*.

➢sample mean

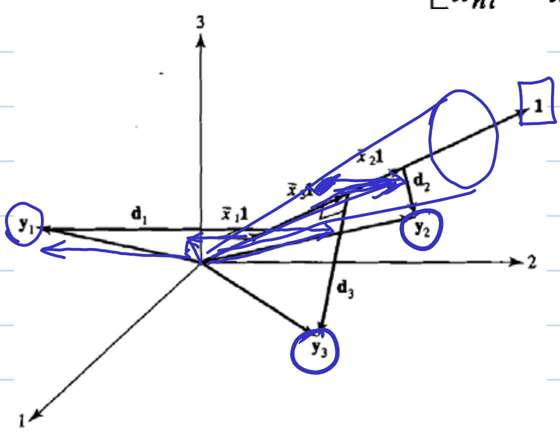
$$\mathbf{y}_i'\left(\frac{1}{\sqrt{n}}\mathbf{1}\right)\underbrace{\frac{1}{\sqrt{n}}\mathbf{1}}_{\text{unit vector}} = \frac{x_{1i}+x_{2i}+\cdots+x_{ni}}{n}\mathbf{1} = \bar{x}_i\mathbf{1}$$

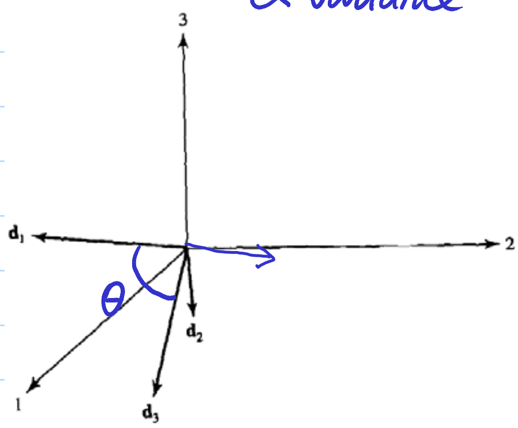
$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \qquad \bar{x}_i\mathbf{1} = \begin{bmatrix} \bar{x}_i \\ \vdots \\ \bar{x}_i \end{bmatrix}$$



➢ sample variance

$$\mathbf{d}_i = \mathbf{y}_i - \bar{x}_i\mathbf{1} = \begin{bmatrix} x_{1i}-\bar{x}_i \\ x_{2i}-\bar{x}_i \\ \vdots \\ x_{ni}-\bar{x}_i \end{bmatrix} \qquad L^2_{\mathbf{d}_i} = \mathbf{d}_i'\mathbf{d}_i = \sum_{j=1}^{n}(x_{ji}-\bar{x}_i)^2 \propto \text{variance}$$





➢ sample covariance and correlation

$$\mathbf{d}_i'\mathbf{d}_k = L_{\mathbf{d}_i}L_{\mathbf{d}_k}\cos(\theta_{ik}) = \sum_{j=1}^{n}(x_{ji}-\bar{x}_i)(x_{jk}-\bar{x}_k) \propto \text{covariance}$$

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

- 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