5. Bivariate and vector random variables

- If we have a collection of random variables $Y_1, Y_2, \ldots, Y_n$ we can gather them together into a vector random variable $\mathbf{Y}$.
- Suppose that, for each $i = 1, \ldots, n$ we have $\mathrm{E}[Y_i] = \mu_i$. Then, we write $\mathrm{E}[\mathbf{Y}] = \boldsymbol{\mu}$ for $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$.
- Now, write $\mathrm{Cov}(Y_i, Y_j) = V_{ij}$ for $i \neq j$ and $\mathrm{Var}(Y_i) = \mathrm{Cov}(Y_i, Y_i) = V_{ij}$. We call $\mathbb{V} = [V_{ij}]_{n \times n}$ the **variance-covariance matrix** for $\mathbf{Y}$.
- We can also call $\mathbb{V}$ the **covariance matrix** or, more simply, just the **variance matrix**. We write $\mathbb{V} = \mathrm{Var}(\mathbf{Y})$.

Example. Let $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ be a vector consisting of n independent random variables, each with mean zero and variance σ^2. This is a common model for **measurement error** on n measurements. We have

$$\mathrm{E}[\boldsymbol{\epsilon} = \mathbf{0}], \qquad \mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbb{I}$$

where $\mathbf{0} = (0, \ldots, 0)$ and $\mathbb{I}$ is the $n \times n$ identity matrix. The off-diagonal entries of $\mathrm{Var}(\boldsymbol{\epsilon})$ are zero since $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$. For measurement error models, we break our usual rule of using upper case letters for random variables.

Example. A population version of the linear model

- First recall the sample version, which is

$$\text{(LM3)} \qquad \mathbf{y} = \mathbb{X}\,\mathbf{b} + \mathbf{e},$$

where $\mathbf{y}$ is the measured response, $\mathbb{X}$ is an $n \times p$ matrix of explanatory variables, $\mathbf{b}$ is chosen by least squares, and $\mathbf{e}$ is the resulting vector of residuals.

- We want to build a random vector $\mathbf{Y}$ that provides a population model for the data $\mathbf{y}$. We write this as

$$\boxed{\text{(LM6)} \qquad \mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}}$$

where $\mathbb{X}$ is the same explanatory matrix as in (LM3), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is an unknown coefficient vector (we don't know the true population coefficient!) and $\boldsymbol{\epsilon}$ is measurement error with $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\mathbb{I}$.

- Our model (LM6) asserts that the process which generated the response data $\mathbf{y}$ was like drawing a random vector $\mathbf{Y}$ consructed using a random measurement error model with known matrix $\mathbb{X}$ for some fixed but unknown value of $\boldsymbol{\beta}$.

Mean and variance of the least squares estimate for $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

- Recall that the main purpose of having a probability model is so that we can investigate the chance variation due to picking the sample.
- Recall that for (LM3), the least squares estimate is $\mathbf{b} = \left(\mathbb{X}^{\mathrm{T}}\mathbb{X}\right)^{-1}\mathbb{X}^{\mathrm{T}}\mathbf{y}$.
- This is a **statistic**, which means a function of the data and not a random variable. We cannot properly talk about the mean and variance of $\mathbf{b}$.
- We can work out the mean and variance of $\left(\mathbb{X}^{\mathrm{T}}\mathbb{X}\right)^{-1}\mathbb{X}^{\mathrm{T}}\mathbf{Y}$, as long as we know how to work out the mean and variance of linear combinations.
- As long as $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is a good **probability model** for the relationship between the response variable $\mathbf{y}$ and the explanatory variable $\mathbb{X}$, calculations done with this model may be useful.

Mean of a linear combination, in matrix form

- The linear property of expectation lets us take expectation through a summation sign, and we get

$$\mathrm{E}\left[\sum_{j=1}^{n} a_{ij} Y_j\right] = \sum_{j=1}^{n} a_{ij} \mathrm{E}[Y_j].$$

- In matrix form, with $\mathbb{A} = [a_{ij}]$, this is

$$\mathrm{E}\left[\mathbb{A}\mathbf{Y}\right] = \mathbb{A}\mathrm{E}[\mathbf{Y}].$$

Example. For $\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$,

$$\mathrm{E}[\mathbf{Y}] = \mathbb{X}\boldsymbol{\beta} + E[\boldsymbol{\epsilon}] = \mathbb{X}\boldsymbol{\beta}$$

Example. $\hat{\boldsymbol{\beta}} = \left(\mathbb{X}^{\mathrm{T}}\mathbb{X}\right)^{-1}\mathbb{X}^{\mathrm{T}}\mathbf{Y}$, we have