Matrices

- **Matrices** are rectangular arrays of numbers
- **dimension** is the number of rows and columns, e.g. 3x2 or 13x3
- We can represent our adsorption MLR by using matrices:

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} + \epsilon_2 \\
&\vdots \\
y_{13} &= \beta_0 + \beta_1 x_{13,1} + \beta_2 x_{13,2} + \epsilon_{13}
\end{aligned}
$$

Gets turned into

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Matrix Formula for MLR

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ 1 & x_{3,1} & x_{3,2} \\ 1 & x_{4,1} & x_{4,2} \\ 1 & x_{5,1} & x_{5,2} \\ 1 & x_{6,1} & x_{6,2} \\ 1 & x_{7,1} & x_{7,2} \\ 1 & x_{8,1} & x_{8,2} \\ 1 & x_{9,1} & x_{9,2} \\ 1 & x_{10,1} & x_{10,2} \\ 1 & x_{11,1} & x_{11,2} \\ 1 & x_{12,1} & x_{12,2} \\ 1 & x_{13,1} & x_{13,2} \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \end{pmatrix}$$

Matrix Operations: Addition and Multiplication

- Matrix operations require **conformable** matrices
- Addition: $\mathbf{A}+\mathbf{B}$
  - Matrices must have the same dimension, i.e. the same number of rows/columns
  - $(i,j)^{th}$ element of $\mathbf{A}+\mathbf{B}$ comes from adding $(i,j)^{th}$ elements of $\mathbf{A}$ and $\mathbf{B}$
  - $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$
- Multiplication: $\mathbf{AB}$
  - $\mathbf{A}$ must have same number of columns as $\mathbf{B}$ has rows
  - $(i,j)^{th}$ element of $\mathbf{AB}$ comes from the dot-product of i^{th} row of $\mathbf{A}$ and j^{th} column of $\mathbf{B}$
  - In general, $\mathbf{AB} \neq \mathbf{BA}$

Transposition and Inverses

- Transposition: $\mathbf{A}^T = \mathbf{A}'$
  - Swaps the rows and the columns of a matrix
  - No restriction on dimension
- Inverting:
  - The **inverse** of a matrix $\mathbf{A}$ is denoted $\mathbf{A}^{-1}$
  - Can only be applied to square matrices
  - Not all square matrices can be inverted
  - $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$
- $\mathbf{I}$?
  - Square and symmetric
  - Diagonal with main diagonal of 1s, all other entries are 0s
  - $\mathbf{I_n}$ has dimension $n\ x\ n$
  - $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ if $\mathbf{I}$ and $\mathbf{A}$ are conformable

Matrix Operation Examples

- Let $\mathbf{A} = \begin{pmatrix} 7 & 5 \\ 5 & 2 \\ 3 & 2 \end{pmatrix}$, and $\mathbf{B} = \begin{pmatrix} 4 & 2 & 1 & 0 \\ 3 & 1 & 1 & 0 \end{pmatrix}$
- We can multiply **AB** since **A** is 3x2 and **B** is 2x4

$$\mathbf{AB} = \begin{pmatrix} 7*4+5*3 & 7*2+5*1 & 7*1+5*1 & 7*0+5*0 \\ 5*4+2*3 & 5*2+2*1 & 5*1+2*1 & 5*0+2*0 \\ 3*4+2*3 & 3*2+2*1 & 3*1+2*1 & 3*0+2*0 \end{pmatrix}$$

$$= \begin{pmatrix} 43 & 19 & 12 & 0 \\ 26 & 12 & 7 & 0 \\ 18 & 8 & 5 & 0 \end{pmatrix}$$

- Neither **A** nor **B** have an inverse, but we can transpose them. For example

$$\mathbf{A}^T = \mathbf{A}' = \begin{pmatrix} 7 & 5 & 3 \\ 5 & 2 & 2 \end{pmatrix}$$

Matrix Operations for MLR

For our system of equations: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

$$\mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} \\ 1 & x_{2,1} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{13,1} & x_{13,2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} \\ \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} \\ \vdots \\ \beta_0 + \beta_1 x_{13,1} + \beta_2 x_{13,2} \end{pmatrix}$$

$$\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{pmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} \\ \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} \\ \vdots \\ \beta_0 + \beta_1 x_{13,1} + \beta_2 x_{13,2} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{13} \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \epsilon_1 \\ \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} + \epsilon_2 \\ \vdots \\ \beta_0 + \beta_1 x_{13,1} + \beta_2 x_{13,2} + \epsilon_{13} \end{pmatrix}$$

Assumptions in MLR

$\boldsymbol{\epsilon} \sim N_n(\mathbf{0},\ \sigma^2\mathbf{I}_n)$

- N_n denotes the n-dimensional multivariate normal distribution
- $\mathbf{0}$ is a vector of zeros - indicating that our proposed model is correct, on average
- $\sigma^2\mathbf{I}_n$ is the variance-covariance matrix of the residuals

$$\mathrm{V}\left[\boldsymbol{\epsilon}\right] = \sigma^2\mathbf{I}_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

  - Diagonal entries are variances: $\mathrm{V}\left[\epsilon_1\right]$, $\mathrm{V}\left[\epsilon_2\right]$ etc.
  - Off-diagonal entries are covariances: $\mathrm{COV}\left[\epsilon_1, \epsilon_2\right]$
  - Zero covariances $\Rightarrow$ Zero correlation.
    - Zero Correlation + Normality $\Rightarrow$ Independence
- Covariance matrix is diagonal in a CRD - not true in other designs

Matrices for Model Estimation

- Need to minimize *SSE* to estimate $\boldsymbol{\beta}$

$$SSE = \hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}} = \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^T \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

- Using calculus to minimize *SSE* results in a system of equations called the **normal equations**

$$\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{y}$$

- Solving the normal equations yields our parameter estimates

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

- If we want to get the predicted responses, $\hat{\mathbf{y}}$, we use

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y}$$

    - $\mathbf{H}$ is called the **hat matrix** because it puts the hats on y values
    - $\mathbf{H}$ has some amazing properties that make many MLR-related computations very easy, e.g. $SSE = \mathbf{y}^T(\mathbf{I} - \mathbf{H})\mathbf{y}$.

Matrices for Inference: Part 1

- Using expectation and variance properties from ST 511 we can show:
  - $\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] = \boldsymbol{\beta}$
  - $\mathrm{V}\left[\hat{\boldsymbol{\beta}}\right] = \mathrm{V}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right] = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$
- Thus:
  - Parameter estimates are unbiased!
  - Variance-Covariance matrix depends on the common variance about the response, σ^2, and the relationship between the predictors (like SLR?)
- Note that $\boldsymbol{\Sigma} = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ is unknowable!
  - $\hat{\boldsymbol{\Sigma}} = MSE\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$
- $\boldsymbol{\Sigma}$ not diagonal $\Rightarrow$ dependent parameter estimates

Matrices for Inference: Part 2

- Sample residuals:

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

- New observation:

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ x_{0,1} \\ x_{0,2} \\ \vdots \\ x_{0,p} \end{pmatrix}$$

- Predicting a Mean Response: $\mathbf{x}_0^T\hat{\boldsymbol{\beta}}$
- Estimated Variance of a Mean Response: $\mathbf{x}_0^T\hat{\boldsymbol{\Sigma}}\mathbf{x}_0$
- Estimated Variance of an Individual Response: $\mathbf{x}_0^T\hat{\boldsymbol{\Sigma}}\mathbf{x}_0 + MSE$

Matrices for Adsorption! - Part 1

Example 5.1

$$\mathbf{y} = \begin{pmatrix} 4 \\ 18 \\ 14 \\ 18 \\ 26 \\ 26 \\ 21 \\ 30 \\ 28 \\ 36 \\ 65 \\ 62 \\ 40 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 13 & 61 \\ 1 & 21 & 175 \\ 1 & 24 & 111 \\ 1 & 23 & 124 \\ 1 & 64 & 130 \\ 1 & 38 & 173 \\ 1 & 33 & 169 \\ 1 & 61 & 169 \\ 1 & 39 & 160 \\ 1 & 71 & 244 \\ 1 & 112 & 257 \\ 1 & 88 & 333 \\ 1 & 54 & 199 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Matrices for Adsorption! - Part 2

Example 5.1

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 13 & 641 & 2305 \\ 641 & 41831 & 133162 \\ 2305 & 133162 & 467669 \end{pmatrix} \qquad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \begin{pmatrix} -7.3507 \\ 0.3490 \\ 0.1127 \end{pmatrix}$$

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 0.633138 & 0.002477 & -0.003826 \\ 0.002477 & 0.000265 & -0.000088 \\ -0.003826 & -0.000088 & 0.000046 \end{pmatrix}$$

$$MSE = \frac{SSE}{n-(p+1)} = \frac{\mathbf{y}^T(\mathbf{I}-\mathbf{H})\mathbf{y}}{13-3} = \frac{191.7897}{10} = 19.17897$$

$$\hat{\boldsymbol{\Sigma}} = MSE\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 12.14294 & 0.04750 & -0.07337 \\ 0.04750 & 0.00508 & -0.00168 \\ -0.07337 & -0.00168 & 0.00088 \end{pmatrix}$$

Inference for Adsorption! - Part 1

Example 5.1

- What is the fitted equation?

$$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = -7.3507 + 0.3490 x_1 + 0.1127 x_2$$

- What is the interpretation for $\hat{\beta}_2$?

  $\hat{\beta}_2 = 0.1127$ is the expected change in adsorption for a one unit increase in extractable iron while holding the amount of extractable aluminum constant

- What is the estimated standard error of $\hat{\beta}_2$?

  $\sqrt{0.00088} = 0.0297$

Inference for Adsorption! - Part 2

Example 5.1

- After accounting for the relationship between extractable aluminum and adsorption, conduct a test to see if extractable aluminum has a significant effect on the adsorption index. (*This is called a Type III test.*)

  $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$ (assuming x_1 is in the model).

  $$t = \frac{\hat{\beta}_2 - 0}{\widehat{SE}\left[\hat{\beta}_2\right]} = \frac{0.1127 - 0}{0.0297} = 3.795$$

  $p - value = 2P(t_{10} \geq 3.795) = 0.0026$

  Reject H_0 since $0.0026 < 0.05$

  There is sufficient sample evidence at the 5% level that extractable iron has a significant linear relationship with adsorption index, even after accounting for extractable aluminum

Inference for Adsorption! - Part 3

Example 5.1

- Estimate the mean adsorption index when soil has an extractable aluminum value of 100 units and extractable iron of 150 units

$$\hat{\mu} = \mathbf{x}_0^T \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 100 & 150 \end{pmatrix} \begin{pmatrix} -7.3507 \\ 0.3490 \\ 0.1127 \end{pmatrix}$$

$$= (1)(-7.3507) + (100)(0.3490) + (150)(0.1127) = 44.454$$

- Provide the standard error for your predicted mean.

$$\mathbf{x}_0^T \hat{\boldsymbol{\Sigma}} \mathbf{x}_0 = \begin{pmatrix} 1 & 100 & 150 \end{pmatrix} \begin{pmatrix} 12.14294 & 0.04750 & -0.07337 \\ 0.04750 & 0.00508 & -0.00168 \\ -0.07337 & -0.00168 & 0.00088 \end{pmatrix} \begin{pmatrix} 1 \\ 100 \\ 150 \end{pmatrix}$$

$$= 19.832$$

Standard Error is $\sqrt{19.832} = 4.453$

Inference for Adsorption! - Part 4

Example 5.1

- Find and interpret a 95% confidence interval for your prediction.

  $\hat{\mu} \pm t_{10,\ 0.025} \hat{SE}[\hat{\mu}] = 44.454 \pm 2.228(4.453)$
  $= 44.454 \pm 9.921 = (34.533,\ 54.375)$

  We are 95% confident that the true mean adsorption index among the population of all soil samples with extractable aluminum of 100 units and extractable iron of 150 units is between 34.5 and 54.4.

- Find the 95% prediction interval.

  $\hat{\mu} \pm t_{10,\ 0.025} \hat{SE}[\hat{y}] = 44.454 \pm 2.228\sqrt{19.832 + 19.17897}$
  $= 44.454 \pm 13.916 = (30.535,\ 58.370)$