The material from this lesson is drawn from "Linear Algebra and its Application", 5th Edition, by Lay, "A First Course in Linear Model Theory" by Ravishanker and Dey, and from Dr. Steve MacEachern's Spring 2014 course "STAT 6860: Foundations of the Linear Model" at The Ohio State University.

I. Linear Algebra + Statistics

Suppose we have a spreadsheet with ten different variables:

- Y: a continuous variable we want to predict
- $X_1,\ldots,X_{10}$: ten independent variables we want to use to predict possible values of Y

Table 1: Spreadsheet of Data

Y	X_1	X_2	$\ldots$	X_{10}
y_1	x_{11}	x_{12}	$\ldots$	$x_{1,10}$
y_2	x_{21}	x_{22}	$\ldots$	$x_{2,10}$
y_3	x_{31}	x_{31}	$\ldots$	$x_{3,10}$

We could write each observation as a separate linear equation, using β to denote unknown coefficient values.

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \cdots + \beta_{10} x_{1,10} \\ y_2 &= \beta_0 + \beta_1 x_{21} + \cdots + \beta_{10} x_{2,10} \\ &\ldots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \cdots + \beta_{10} x_{n,10} \end{aligned}$$

As we did last week, rather than representing these all as individual equations, let's put them into matrix form.

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,10} \\ x_{21} & x_{22} & \cdots & x_{2,10} \\ \ldots & & & \\ x_{n1} & x_{n2} & \cdots & x_{n,10} \end{bmatrix}$$

Our goal will be to find the values of β that best parameterize the relationship between $\mathbf{Y}$ and $\mathbf{X}$.

II. Linear Combinations

We say that W is a **linear combination** of Z if we can write W as $a + bZ$, where a and b are some real numbers.

Think about a particular vector (or variable) as contributing some information toward our solution of a problem. If one variable X_1 is linearly related to another variable X_2, then having both variables included in our model provides no new information about Y. Including either X_1 or X_2 should be sufficient.

What happened when we tried to include both male and female in a model?

We say that $\mathbf{Y}$ is a linear combination of $\mathbf{X}$ because we believe that $\mathbf{Y}$ can be expressed as the sum of columns of $\mathbf{X}$ multiplied by some scalar constant. (In this case, the scalar constants for each vector will be given by the elements of $\boldsymbol{\beta}$.)

Assume we have matrices $\mathbf{X}$, $\mathbf{Y}$, and $\boldsymbol{\beta}$ as described above. The ordinary least-squares estimate (or the estimates for $\boldsymbol{\beta}$ that we would get from minimizing mean squared error) would be given by:

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

However, this is only true if $\mathbf{X}$ is **invertible**.

III. Invertibility

Whether or not a matrix is invertible is of paramount importance in linear algebra and its application. If a design matrix $\mathbf{X}$ is not invertible, then the estimators we get for $\boldsymbol{\beta}$ will not be the *best linear unbiased estimators.*

But what does it mean for $\mathbf{X}$ to be invertible?

It simply means that the matrix $\mathbf{X}$ can be inverted.

Is there a better way to detect for invertibility?

Indeed, there is.

One thing left off of this particular link is that if the determinant of the matrix $\mathbf{X}$ is equal to zero, then the matrix is not invertible.