I. Linear Model Recap & Extensions

Recall that our linear model is given by $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and that our goal is to find values of $\boldsymbol{\beta}$ that best define the relationship between $\mathbf{Y}$ and $\mathbf{X}$.

- $\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{y})$
- $\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}\mathbf{y})$
- $\hat{\boldsymbol{\beta}}_{i+1} = (\mathbf{X}^T\mathbf{W}_i\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}_i\mathbf{y}) \Rightarrow \hat{\boldsymbol{\beta}}_{IRLS}$
- $\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\mathbf{X})^{-}(\mathbf{X}^T\mathbf{y})$

II. "Best" Estimators

We've talked a lot about how to "best" estimate the relationship between X and Y. Specifically, we seek to minimize the sum of the squared residuals and know that we get the estimators above. However, there are other benefits we haven't discussed. What does "best" really mean?

In particular, the Gauss-Markov Theorem - arguably the most important theorem after the Central Limit Theorem - says under certain conditions that the estimators above are **BLUE**.

BLUE means:

- Best
- Linear
- Unbiased
- Estimator

In this case, **linear** means we use a linear model to estimate or predict the quantity of interest.

In this case, **estimator** means we have a formula to estimate or predict the quantity of interest.

In this case, **unbiased** means that, on average, our estimations or predictions will be correct.

In this case, **best** means that, of all possible linear unbiased estimators, the model we construct has the smallest variance.

- If we're willing to have biased estimators, we may find estimators of significantly lower variance... but unbiased is a quality we usually like to have when building models.

What are the certain conditions for our model to be the best linear unbiased estimator?

- X and Y are linearly related to one another.
- Our residuals are independent of one another. (It is equivalent to say that our observations themselves are independent of one another.)
- Our residuals have constant variance.
- Our independent variables are independent of one another.

Recall our assumptions for multiple linear regression:

- **Linearity:** X and Y are linearly related to one another.
- **Independence of Residuals:** Our residuals are independent of one another. (It is equivalent to say that our observations themselves are independent of one another.)
- **Normality of Residuals:** Our residuals are Normally distributed with mean 0.
- **Equality of Variance:** Our residuals have constant variance.
- **Independence of Independent Variarbles:** Our independent variables are independent of one another.

Note that we don't *need* Normality of residuals for our model to be BLUE... so what *does* Normality give us?

Honestly, we just get nicer results about the distribution of the variance of the model and coefficients and we can conduct more parametric inference as a result.

III. Violating Assumptions

When the conditions for BLUE above are violated, we are no longer guaranteed that our model is the best linear unbiased estimator for Y. Usually Python won't trigger alarms if certain conditions aren't met, so it's on us to ensure that they are. (Luckily, the linear model is pretty robust so as long as we don't have extreme violations of these assumptions, we *should* be alright.)

In some cases, though, Python may trigger warnings.

- `Error: This might indicate that that there are strong multicollinearity problems or that the design matrix is singular.`
- `Error: Parameter estimates may be unstable.`

These indicate that our variables are not sufficiently independent of one another. Either we have a perfect *linear combination* within our X variables or we are close enough to a perfect linear combination that, when the computer attempts to invert $\mathbf{X}^T\mathbf{X}$, Python recognizes there's some potential for error. In this case, explore and consider removing variables until you're confident that there isn't an issue.

What if we could devise a method to force our variables to become independent?

That's PCA, and we'll get into that next week.

- PCA relies on linear algebra terms "eigenvalues" and "eigenvectors." We briefly discussed this last week when we mentioned the spectral decomposition/eigendecomposition of a matrix. This was $\mathbf{PDP}^{-1}$, and we used it to very efficiently find the value of A^{100}.

IV. Eigenvalues and Eigenvectors

For *some* matrices $\mathbf{A}$, there are **nonzero** vectors $\mathbf{x}$ such that $\mathbf{Ax}$ is just a scalar multiple of $\mathbf{x}$. (We most frequently write this $\mathbf{Ax} = \lambda\mathbf{x}$.) The vectors $\mathbf{x}$ and the scalars λ for which this equation holds true are called **eigenvectors** and **eigenvalues**.

Consider the matrix $\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$ and the vectors $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

$$\begin{aligned}
\mathbf{Av} &= \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} (3 \times 2) + (-2 \times 1) \\ (1 \times 2) + (0 \times 1) \end{bmatrix} \\
&= \begin{bmatrix} 4 \\ 2 \end{bmatrix} \\
&= 2\mathbf{v}
\end{aligned}$$

Because $\mathbf{Av} = \lambda\mathbf{v}$, the vector $\mathbf{v}$ is an eigenvector of the matrix $\mathbf{A}$. The corresponding eigenvalue is $\lambda = 2$. (We say that $\lambda = 2$ is an eigenvalue of $\mathbf{A}$.)

$$\begin{aligned}
\mathbf{Au} &= \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} (3 \times -1) + (-2 \times 1) \\ (1 \times -1) + (0 \times 1) \end{bmatrix} \\
&= \begin{bmatrix} -5 \\ -1 \end{bmatrix}
\end{aligned}$$

Because $\mathbf{Au} \neq \lambda\mathbf{u}$, the vector $\mathbf{u}$ is NOT an eigenvector of the matrix $\mathbf{A}$. There is no eigenvalue here.

Consider the equation $\mathbf{Ax} = \lambda\mathbf{x}$:

$$\begin{aligned} \mathbf{Ax} &= \lambda\mathbf{x} \\ \Rightarrow \mathbf{Ax} - \lambda\mathbf{x} &= \mathbf{0} \\ \Rightarrow \mathbf{Ax} - \lambda\mathbf{Ix} &= \mathbf{0} \\ \Rightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} &= \mathbf{0} \end{aligned}$$

There is a way where we can identify all eigenvectors $\lambda_1, \ldots, \lambda_k$ of our matrix $\mathbf{A}$.

A scalar λ is an eigenvalue of an $n \times n$ matrix $\mathbf{A}$ if and only if λ satisfies the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, where $\det(\cdot)$ indicates the determinant.

While we will usually leave calculating the determinant to a CAS (computer algebra system), there are two types of matrices we will explore calculating the determinant by hand:

- A 2×2 matrix of the form $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has determinant $\det(\mathbf{A}) = ad - bc$.
- A triangular matrix $\mathbf{A}$ will have determinant $\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii}$. (That is, the determinant will equal the product of the elements on the diagonal.)

Let's find all eigenvalues of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= 0 \\ \Rightarrow 0 &= \det(\mathbf{A} - \lambda\mathbf{I}) \\ &= \det\left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}\right) \\ &= (2-\lambda)(-6-\lambda) - (3)(3) \\ &= -12 + \lambda^2 + 4\lambda - 9 \\ &= \lambda^2 + 4\lambda - 21 \\ &= (\lambda + 7)(\lambda - 3) \\ \Rightarrow \lambda &= -7, +3 \end{aligned}$$

Thus, $\lambda = -7$ and $\lambda = 3$ are the two eigenvalues of $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

Now that we have the eigenvalues of $\mathbf{A}$, we can find their corresponding eigenvectors. Recall that we showed above that $\mathbf{Ax} = \lambda\mathbf{x}$ implies that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. Let's find the eigenvector associated with

$\lambda = -7$.

$$\begin{aligned}
\mathbf{0} &= (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} \\
\Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} + 7\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \left(\begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} + \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}\right)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\Rightarrow \begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \begin{bmatrix} 9x_1 + 3x_2 \\ 3x_1 + 1x_2 \end{bmatrix} \\
\Rightarrow \begin{bmatrix} -3x_2 \\ -1x_2 \end{bmatrix} &= \begin{bmatrix} 9x_1 \\ 3x_1 \end{bmatrix} \\
\Rightarrow -3x_1 &= x_2 \\
\Rightarrow \mathbf{x} &= \begin{bmatrix} x_1 \\ -3x_1 \end{bmatrix}
\end{aligned}$$

Taking $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ and plugging -7 in for λ, we see that $\mathbf{x}^T = [x_1 \quad -3x_1]$ is an eigenvector for $\mathbf{A}$ that corresponds to the eigenvalue $\lambda = -7$. (We could plug in any value for x_1 here.)

Let's check to see if this actually works!

$$\begin{aligned}
\mathbf{Ax} &= \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}\begin{bmatrix} x_1 \\ -3x_1 \end{bmatrix} \\
&= \begin{bmatrix} 2x_1 - 9x_1 \\ 3x_1 + 18x_1 \end{bmatrix} \\
&= \begin{bmatrix} -7x_1 \\ 21x_1 \end{bmatrix} \\
&= -7\begin{bmatrix} x_1 \\ -3x_1 \end{bmatrix} \\
&= \lambda\mathbf{x}
\end{aligned}$$

This clearly does work.

On your own, find an eigenvector $\mathbf{x}$ associated with $\lambda = 3$, then show that your chosen $\mathbf{x}$ satisfies $\mathbf{Ax} = 3\mathbf{x}$.

V. Applications of Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors have many different applications. While we will go through a few in particular in later weeks, we mention some here:

- Eigenvectors and eigenvalues are used in the diagonalization of matrices, which has incredible computational benefits including applications in solving systems of linear equations and Monte Carlo simulations.
- Principal component analysis is a method by which potentially linearly dependent vectors can be linearly transformed and combined into vectors which are orthogonal to one another.
- Systems of differential equations, a set of equations relating functions to their derivatives, can be solved using eigenvectors and eigenvalues.
- When using matrices to describe the spread of infectious diseases, the basic reproductive number R_0 (indicating the average number of people that a standard infectious person will infect) corresponds to the largest eigenvalue of the "next generation matrix."
- For a matrix $\mathbf{A}$, its determinant can be calculated as $\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$. (Note that if any of the eigenvalues are zero, then the determinant is zero, implying that $\mathbf{A}$ is not invertible.)