INTRODUCTION
In this section, we apply what we have learned about vectors and matrices to a statistics problem. We learn how to find the equation of the line that most nearly contains a collection of points that are not lined up. Fitting lines and curves to data points is known as curve fitting.

The discussion begins with a problem that seems unrelated – trying to solve a system of equations that has no solution.

For example, consider the system of equations,

\[
\begin{align*}
3x &= 5 \\
2x &= 2.
\end{align*}
\]

It is not hard to see that the system has no solution, because the equations don’t agree. Sometimes we call a system like this over-determined. It has more equations than unknowns. Over-determined systems often do not have solutions.

Geometry can demonstrate the fact that there is no solution. The vector form of the system is

\[
\begin{pmatrix} 3 \\ 2 \end{pmatrix} x = \begin{pmatrix} 5 \\ 2 \end{pmatrix}.
\]

If \( a = \begin{pmatrix} 3 \\ 2 \end{pmatrix} \) and \( b = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \), the vector form is

\[ ax = b. \]

Since \( x \) is a scalar, the system seeks a way to scale (stretch) vector \( a \) so that the result is equal to vector \( b \). Remember, multiplying by a scalar can only stretch a vector, but never change its direction. The vectors \( a \) and \( b \) are plotted below.

Since \( a \) and \( b \) are not parallel, there is no way to scale \( a \) so that the result is \( b \). Thus, the system of equations has no solution.
But, what if we were to scale $\mathbf{a}$ so that the difference between $\mathbf{a}$ and $\mathbf{b}$ was as small as possible? The image below shows a scaled version of $\mathbf{a}$, denoted $\mathbf{a}_x$, as well as the difference between $\mathbf{a}_x$ and $\mathbf{b}$, denoted as $\mathbf{e}$.

1. Think about the vectors $\mathbf{a}_x$, $\mathbf{e}$, and $\mathbf{b}$. Two of these add up to the other. Write this relationship as an equation below.

2. When you are sure that your equation correctly relates the vectors mentioned above, solve the equation for the vector, $\mathbf{e}$.

   \[ \mathbf{e} = \underline{\quad} \]

The name of the vector $\mathbf{e}$ stands for error. It is the difference between $\mathbf{a}_x$ and $\mathbf{b}$. While it is impossible for the equation, $\mathbf{a}_x = \mathbf{b}$, to be true, we can still try to make the error vector, $\mathbf{e}$, as small as possible.

3. The image below shows many possible scalings, $\mathbf{a}_x$, of the vector $\mathbf{a}$, along with the corresponding error vectors, $\mathbf{e}$. We wish to choose $x$ so that the length of $\mathbf{e}$ is as small as possible. Circle the letter $\mathbf{e}$ corresponding to its smallest representation.
4. What is the angle between the shortest possible vector and a? What is the word used to describe the relationship between such vectors?

5. What is the value of the dot product between the shortest possible vector and a?

6. Write an equation that states the value of this dot product.

\[ \mathbf{a} \cdot \mathbf{e} = ____ \]

7. Rewrite the equation above, replacing \( \mathbf{e} \) with the expression developed in question 2.

8. Dot products have a distributive property just like multiplication. Distribute through the parentheses in the above equation, and move the negative dot product to the other side.

9. Use the definitions of a and b above to compute the two dot products in the equation in question 8. Rewrite the equation with the resulting values, and solve for x.

The solution to the over-determined system that you just found is called a least-squares solution. We say least-squares because it minimizes the (square) of the length of the error vector, \( \mathbf{e} \). The squared length is considered because it is simpler than the length itself. (Remember, if \( \mathbf{e} = \begin{pmatrix} r \\ s \end{pmatrix} \), then the length of \( \mathbf{e} \) is \( |\mathbf{e}| = \sqrt{r^2 + s^2} \). The square of the length is simpler, \( |\mathbf{e}|^2 = r^2 + s^2 \), and this is what is minimized in a least-squares solution.)

10. Recall that if we wish to represent a dot product such as \( \mathbf{v} \cdot \mathbf{w} \) as a matrix multiplication, we have to take the transpose of \( \mathbf{v} \) to make it a row vector: \( \mathbf{v}^T \cdot \mathbf{w} \). Rewrite the equation from question 9 with the dot products expressed as matrix multiplications.
CHECK IT OUT!

If \( Ax = b \) is the vector form of an over-determined linear system of equations, then the least-squares solution can be found by solving

\[
A^\top Ax = A^\top b.
\]

Be careful to multiply the matrices in the order given above because order matters in matrix multiplication.

YOUR TURN!

Consider the system of equations

\[
\begin{align*}
x + 2y &= 3 \\
2x - 3y &= -1 \\
5x + y &= 5.
\end{align*}
\]

11. Rewrite the system in vector form, \( Ax = b \). The coefficients of \( x \) and \( y \) form the matrix \( A \), and the unknowns \( x \) and \( y \) themselves make up the column vector, \( x \). The constants on the right sides make up the column vector, \( b \).

12. The least squares solution of the system above is determined by solving \( A^\top Ax = A^\top b \).
   a. Compute \( A^\top A \).

   b. Compute \( A^\top b \)

   c. Rewrite the vector form \( A^\top Ax = A^\top b \) in vector form, then again as a regular system of equations with two equations and two unknowns.
d. Solve the system.

13. An important tool in statistics is curve fitting. Suppose you want to find the line which most nearly contains the points (-2, 1), (0, 4), and (2, 9). The equation of the line would look like $y = mx + b$.
   a. Graph these points below.

   ![Graph of points]

   b. Substitute each point into the equation, but write the equation as $mx + b = y$.

   $$
   \begin{align*}
   -2m + b &= 1 \\
   0m + b &= 4 \\
   2m + b &= 9
   \end{align*}
   $$

   c. You have just created a system of equations. Does this system have a solution? Explain your answer.
d. Write your system of equations in vector form \((Ax = b)\).

e. The least squares solution is found by left-multiplying the equation by \(A^T\) to get \(A^TAx = A^Tb\). Compute \(A^TA\).

f. Compute \(A^Tb\).

g. Rewrite the system \(A^TAx = A^Tb\) in vector form, and then as a system with two equations and two unknowns, then solve.

h. You now know the slope and \(y\)-intercept for the least squares line (also called the least squares regression line), \(y = mx + b\). Give this equation of the line below, and graph it on the axes from question 13a.