Orthogonal ?

1. Show that the following vectors are pairwise orthogonal by computing dot products.

   a) $\mathbf{1}$

   b) $\mathbf{x} - \bar{x}\mathbf{1}$

   ☆ c) $\mathbf{y} - \bar{y}\mathbf{1} - b_{\mathbf{y}\|\mathbf{x}_{\perp\mathbf{1}}}(\mathbf{x} - \bar{x}\mathbf{1})$

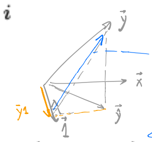
$\vec{e} = \vec{y} - b\,\vec{x}$ (centerized)

Problem 1

1.

$$\mathbf{1} \cdot (\boldsymbol{x} - \bar{x}\mathbf{1}) = \sum_i x_i - \sum_i \bar{x} = \sum_i x_i - n\bar{x} = \sum_i x_i - \sum_i x_i = 0$$

2.



$(y - \bar{y}1)$ = the perpendicular part of $\vec{y}$ on $\vec{1}$ $= y_{\perp 1}$

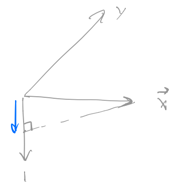
$$\mathbf{1} \cdot (\boldsymbol{y} - \bar{y}\mathbf{1} - b_{y||(x\perp 1)}(\boldsymbol{x} - \bar{x}\mathbf{1})) = \mathbf{1} \cdot \underbrace{(\boldsymbol{y} - \bar{y}\mathbf{1})}_{\text{perpendicular}} - \mathbf{1} \cdot b_{y||(x\perp 1)}(\boldsymbol{x} - \bar{x}\mathbf{1}))$$

$$= 0 - \underbrace{\boxed{b_{y||(x\perp 1)}}}_{\text{a constant}}\mathbf{1} \cdot \underbrace{(\boldsymbol{x} - \bar{x}\mathbf{1})}_{\text{= similarly, the perpendicular part of } \vec{x} \text{ on } \vec{1} \;=\; x_{\perp 1}})$$

$$= 0$$

☆ or, $= 1 \cdot \vec{e}$

$= 0$



3.

$$\begin{aligned}
(\mathbf{x} - \bar{x}\mathbf{1}) \bullet \mathbf{e} &= (\mathbf{x} - \bar{x}\mathbf{1}) \bullet (\mathbf{y} - \bar{y}\mathbf{1} - b_{y\|(x\perp 1)}(\mathbf{x} - \bar{x}\mathbf{1})) \\
&= (\mathbf{x} - \bar{x}\mathbf{1}) \bullet (\mathbf{y} - \bar{y}\mathbf{1}) \\
&\quad - b_{y\|(x\perp 1)}(\mathbf{x} - \bar{x}\mathbf{1})) \bullet (\mathbf{x} - \bar{x}\mathbf{1})
\end{aligned}$$

Note that

$$(\mathbf{x} - \bar{x}\mathbf{1}) \bullet \underbrace{(\mathbf{y} - \bar{y}\mathbf{1})}_{= b\cdot(x - \bar{x}1)} = b_{y\|(x\perp 1)} \parallel \mathbf{x} - \bar{x}\mathbf{1} \parallel^2$$

And that

$$(\mathbf{x} - \bar{x}\mathbf{1}) \bullet b_{y\|(x\perp 1)}(\mathbf{x} - \bar{x}\mathbf{1}) = b_{y\|(x\perp 1)} \parallel \mathbf{x} - \bar{x}\mathbf{1} \parallel^2$$

Therefore,

$$(\mathbf{x} - \bar{x}\mathbf{1}) \bullet \mathbf{e} = 0$$

Derive Properties

2. Use the results from the previous problem to establish the following properties of the residual and fitted value vectors for simple linear regression:

   a) $\bar{\mathbf{e}} = 0$

   b) $\hat{\mathbf{y}} \bullet \mathbf{e} = 0$

   c) $\mathbf{x} \bullet \mathbf{e} = 0$

   d) $\|\mathbf{y} - \bar{y}\mathbf{1}\|^2 = \|\mathbf{e}\|^2 + \|\hat{\mathbf{y}} - \bar{y}\mathbf{1}\|^2$

Problem 2

1. This follows directly from the normal equations. Explicitly, note

$$\begin{aligned}\bar{e} &= \frac{1}{n}\mathbf{1} \cdot \boldsymbol{e} \\ &= \mathbf{1} \cdot (\boldsymbol{y} - \bar{y}\mathbf{1} - b_{y||(x\perp 1)}(\boldsymbol{x} - \bar{x}\mathbf{1})) \\ &= 0\end{aligned}$$

   where a (the intercept) is 0 because $\boldsymbol{y}$ and $\boldsymbol{x}$ have both been centered.

2. $\hat{\mathbf{y}} \bullet \mathbf{e} = 0$

$$\begin{aligned}\hat{\mathbf{y}} \bullet \mathbf{e} &= (\bar{y}\mathbf{1} + b_{y\|(x\perp 1)}(\mathbf{x} - \bar{x}\mathbf{1})) \bullet \mathbf{e} \\ &= \bar{y}\mathbf{1} \bullet \mathbf{e} + b_{y\|(x\perp 1)}(\mathbf{x} - \bar{x}\mathbf{1}) \bullet \mathbf{e} \\ &= 0 \qquad \textit{by (ii) and (iii) above}\end{aligned}$$

3. $\mathbf{x} \bullet \mathbf{e} = 0$

   Note that from (ii) above we know that $\mathbf{1}$ is orthogonal to $\mathbf{e}$ so we can re-express $\mathbf{x} \bullet \mathbf{e}$ as follows:

$$\mathbf{x} \bullet \mathbf{e} = (\mathbf{x} - \bar{x}\mathbf{1}) \bullet \mathbf{e}$$

   And the righthand side of the equation was shown to be 0 in (iii) above.

4. $|\mathbf{y} - \bar{y}\mathbf{1}|^2 = |\mathbf{e}|^2 + |\hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2$

   Add and subtract $\hat{\mathbf{y}}$ within $|\mathbf{y} - \bar{y}\mathbf{1}|^2$ to get:

$$\begin{aligned}|\mathbf{y} - \bar{y}\mathbf{1}|^2 &= |\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2 \\ &= |\mathbf{e}|^2 + |\hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2 + 2\mathbf{e} \bullet (\hat{\mathbf{y}} - \bar{y}\mathbf{1})\end{aligned}$$

The cross-product term is 0 because by (b) we have $\mathbf{e} \bullet \hat{\mathbf{y}} = 0$, and by (a) we have $\mathbf{e} \bullet \mathbf{1} = 0$.

Regression = Projeation

3. Consider the least squares fit:

$$\min_c \sum_i (Y_i - cX_i)^2$$

without intercept ($\vec{1}$)

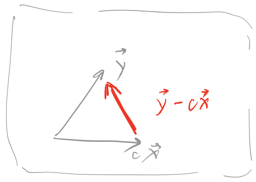
1. Determine $\hat{c}$ that minimizes the above sum of squares.

   Re-express the minimization as

$$\min_c \| \mathbf{y} - c\mathbf{x} \|^2$$

(the Norm of) $\vec{y} - c\vec{x}$

MIN



   Therefore the minimizer is the projection of $\mathbf{y}$ onto the span of $\mathbf{x}$, which is

When $(\vec{y} - c\vec{x})$ is perpendicular to $(c\vec{x})$

the projection of $\vec{y}$ on $\vec{x}$ $=$ $c\vec{x}$

$$\hat{c} = b_{y\|x} = \frac{\mathbf{y} \bullet \mathbf{x}}{\mathbf{x} \bullet \mathbf{x}}$$

i.e. $\frac{\vec{y} \cdot \vec{x}}{\vec{x} \cdot \vec{x}} \cdot \vec{x} = c\vec{x}$

2. Do the properties of the simple linear regression from problem 2 still hold? Prove or disprove. And explain in words why your answer makes sense geometrically.

We begin with the geometric explanation: The residuals from the fit need not be orthogonal to $\mathbf{1}$ because the projection is on $span\{\mathbf{x}\}$ and $\mathbf{1}$ is not in the span (except under certain conditions on $\mathbf{x}$). Therefore, 2(a) does not necessarily hold.

Further, the properties of projection imply that the residuals are orthogonal to the fitted values. This implies 2(b).

Additionally, since the fitted values are in the $span\{\mathbf{x}\}$, the residuals are also orthogonal to $\mathbf{x}$. This implies that 2(c) holds.

Finally, since the residuals are not necessarily orthogonal to $\mathbf{1}$, the cross-product term in the proof of (d) is not necessarily 0 so the variability of $\mathbf{y}$ about the mean does not necessarily equal the sum of the variability in the residuals and the variability of the fitted values about $\bar{y}$. Therefore 2(d) does not hold in this situation.

We can establish these results using the definition of $\hat{c}$ and dot products as follows.

The residuals do not necessarily have an average 0 because:

$$\begin{aligned}
\bar{e} &= \frac{1}{n}\mathbf{e} \bullet \mathbf{1} \\
&= \frac{1}{n}(\mathbf{y} - b_{y\|x}\mathbf{x}) \bullet \mathbf{1} \\
&= \bar{y} - b_{y\|x}\bar{x} \\
&= \bar{y} - \frac{\mathbf{y} \bullet \mathbf{x}}{\mathbf{x} \bullet \mathbf{x}}\bar{x}
\end{aligned}$$

This term can be 0, e.g., when $\bar{y}$ and $\bar{x}$ are 0, but not necessarily.

For 2(b),

$$\begin{aligned}
\hat{\mathbf{y}} \bullet \mathbf{e} &= \hat{\mathbf{y}} \bullet (\mathbf{y} - \hat{\mathbf{y}}) \\
&= b_{y\|x}\mathbf{x} \bullet (\mathbf{y} - b_{y\|x}\mathbf{x}) \\
&= b_{y\|x}^2 \mathbf{x} \bullet \mathbf{x} - b_{y\|x}^2 \mathbf{x} \bullet \mathbf{x} \\
&= 0
\end{aligned}$$

Next 2(c) follows from 2(b) since $\hat{\mathbf{y}} = \hat{c}\mathbf{x}$.

And 2(d) does not necessarily hold because from our work with 2(a), we have that $\mathbf{e} \bullet \mathbf{1}$ is not necessarily 0. This implies that the cross-product term need not be 0, i.e.,

$$\begin{aligned}
|\mathbf{y} - \bar{y}\mathbf{1}|^2 &= |\mathbf{y} - \hat{\mathbf{y}} + \hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2 \\
&= |\mathbf{e}|^2 + |\hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2 + 2\mathbf{e} \bullet (\hat{\mathbf{y}} - \bar{y}\mathbf{1}) \\
&= |\mathbf{e}|^2 + |\hat{\mathbf{y}} - \bar{y}\mathbf{1}|^2 - 2\bar{y}\mathbf{e} \bullet \mathbf{1}
\end{aligned}$$

What can regression model do ?

d) Do you think the models you have fit in this exercise are most useful for summarizing an association, for prediction, or for causal inference? Explain your reasoning.

######### d ############

- these models are best for summarizing an association.
- it is not clear what the prediction context would be. Furthermore, given a new body of work the parameter a is very likely to be different (suppose for example that there are 10 times as many words in the new volume), thus the prediction of the count of words would likely not be very good given new data.
- it is not clear what causal inference would mean here. If i change x (the rank) does y (the frequency) increase? Of course, they are directly related to each other. It is also true that if i change y, x would change. Causality does not seem well defined and we can't really say that b is the effect of a change in rank on frequency.