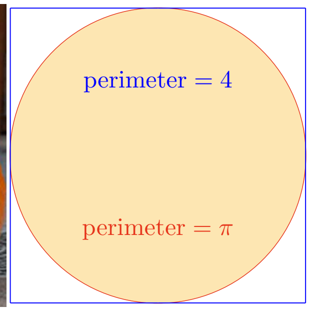
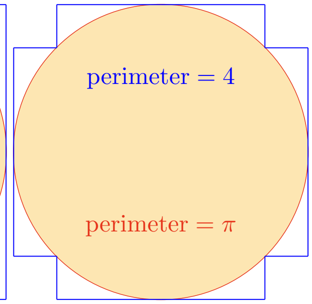
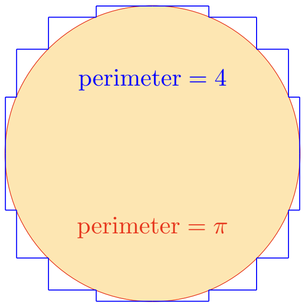
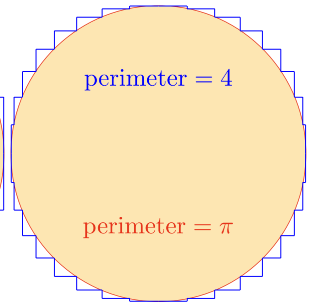
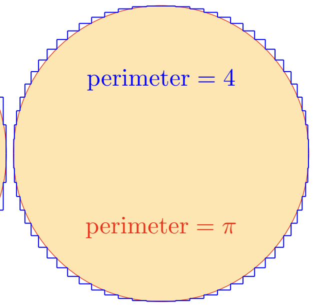
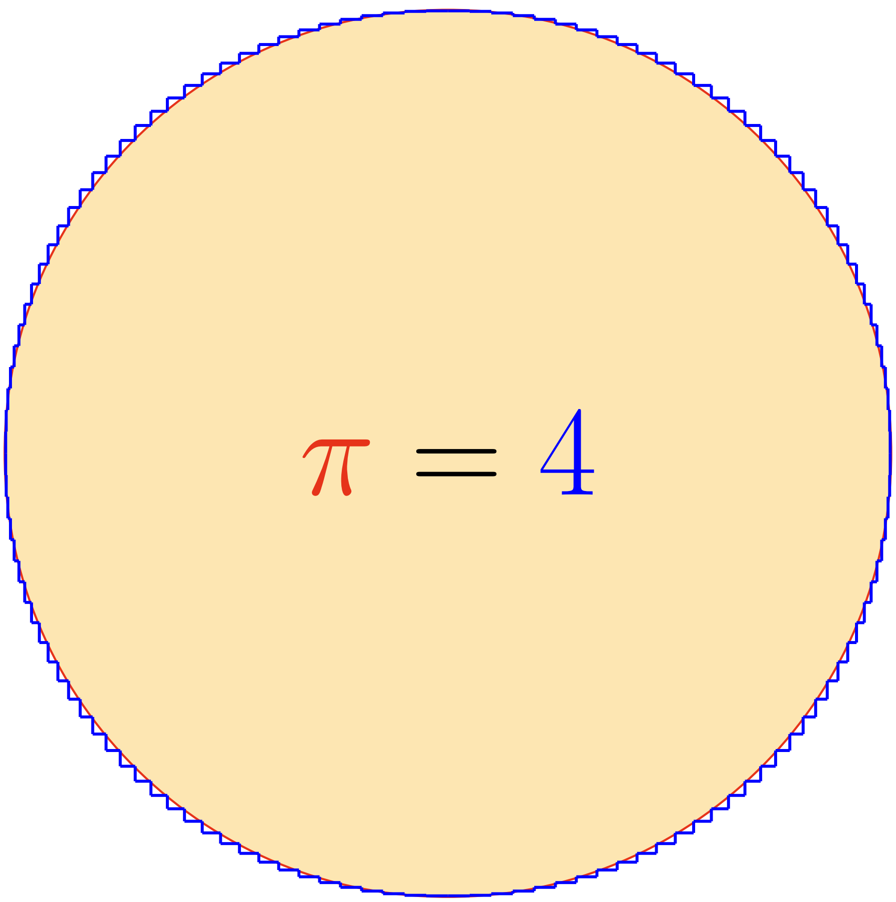












Happy Halloween!

Armin Straub
astraub@illinois.edu

Review

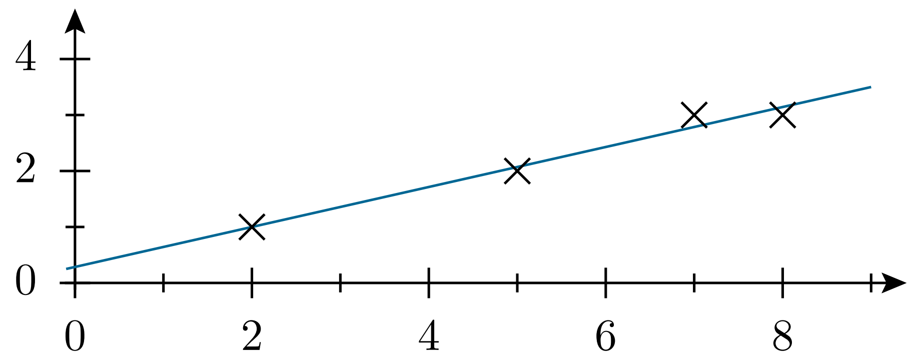
- $\hat{\boldsymbol{x}}$ is a **least squares solution** of the system $A\boldsymbol{x} = \boldsymbol{b}$

  $\Longleftrightarrow$ $\hat{\boldsymbol{x}}$ is such that $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible

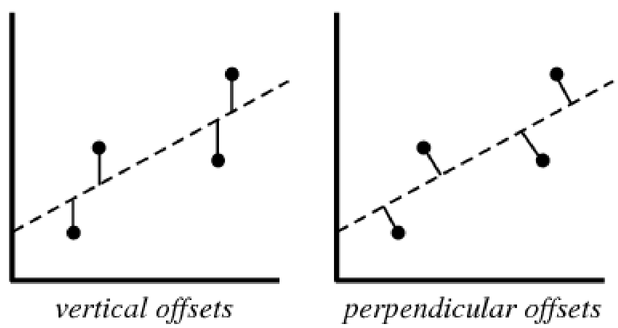
  $\overset{\text{FTLA}}{\Longleftrightarrow}$ $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ (the **normal equations**)

Application: least squares lines

Example 1. Find β_1, β_2 such that the line $y = \beta_1 + \beta_2 x$ best fits the data points $(2,1)$, $(5,2), (7,3), (8,3)$.



Comment. As usual in practice, we are minimizing the (sum of squares of the) vertical offsets:



http://mathworld.wolfram.com/LeastSquaresFitting.html

Solution. The equations $y_i = \beta_1 + \beta_2 x_i$ in matrix form:

Armin Straub
astraub@illinois.edu

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector } \boldsymbol{y}}}$$

Here, we need to find a least squares solution to

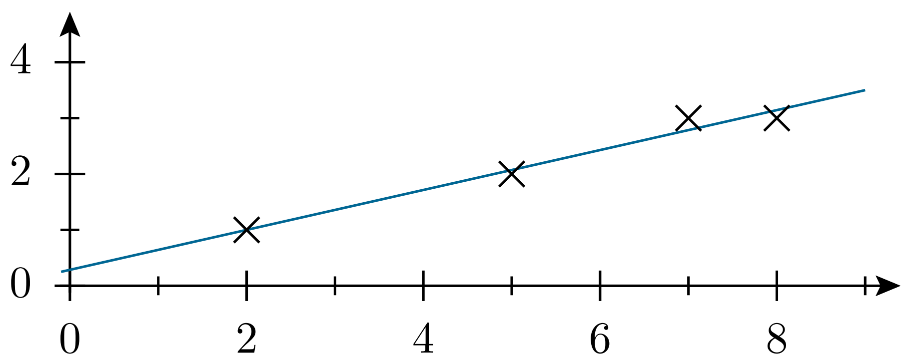
$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

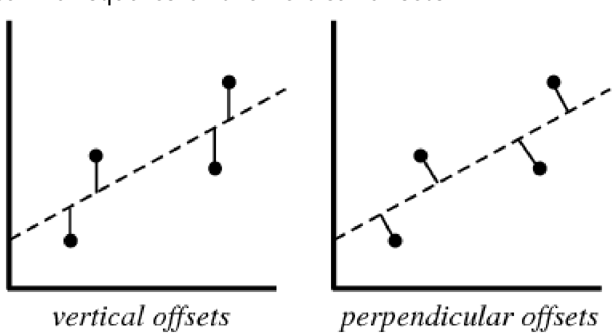
Solving $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \hat{\beta} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Hence, the least squares line is $y = \frac{2}{7} + \frac{5}{14} x$.



How well does the line fit the data $(2,1), (5,2), (7,3), (8,3)$?

How small is the sum of squares of the vertical offsets?



Armin Straub
astraub@illinois.edu

- **residual sum of squares**: $\mathrm{SS}_{\mathrm{res}} = \sum \underbrace{(y_i - (\beta_1 + \beta_2 x_i))}_{\text{error at } (x_i, y_i)}^2$

  The choice of β_1, β_2 from least squares, makes $\mathrm{SS}_{\mathrm{res}}$ as small as possible.

- **total sum of squares**: $\mathrm{SS}_{\mathrm{tot}} = \sum (y_i - \bar{y})^2$,

  where $\bar{y} = \frac{1}{n} \sum y_i$ is the mean of the observed data

- **coefficient of determination**: $R^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}}$

  General rule: the closer R^2 is to 1, the better the regression line fits the data.

  Here, $\bar{y} = 9/4$: $(2,1), (5,2), (7,3), (8,3)$

$$R^2 = 1 - \frac{\left(1 - \left(\frac{2}{7} + \frac{5}{14} 2\right)\right)^2 + \left(2 - \left(\frac{2}{7} + \frac{5}{14} 5\right)\right)^2 + \left(3 - \left(\frac{2}{7} + \frac{5}{14} 7\right)\right)^2 + \left(3 - \left(\frac{2}{7} + \frac{5}{14} 8\right)\right)^2}{\left(1 - \frac{9}{4}\right)^2 + \left(2 - \frac{9}{4}\right)^2 + \left(3 - \frac{9}{4}\right)^2 + \left(3 - \frac{9}{4}\right)^2}$$

$$= 1 - \frac{0.075}{2.75} = 0.974$$

  very close to $1 \Longrightarrow$ good fit

Other curves

We can also fit the experimental data (x_i, y_i) using other curves.

Example 2. $y_i \approx \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ with parameters $\beta_1, \beta_2, \beta_3$.

The equations $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$ in matrix form:

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$$

Given data (x_i, y_i), we then find the least squares solution to $X\beta = \boldsymbol{y}$.

Multiple linear regression

> *In statistics, **linear regression** is an approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables.*
>
> *The case of one explanatory variable is called simple linear regression.*
>
> *For more than one explanatory variable, the process is called multiple linear regression.*
>
> *http://en.wikipedia.org/wiki/Linear_regression*

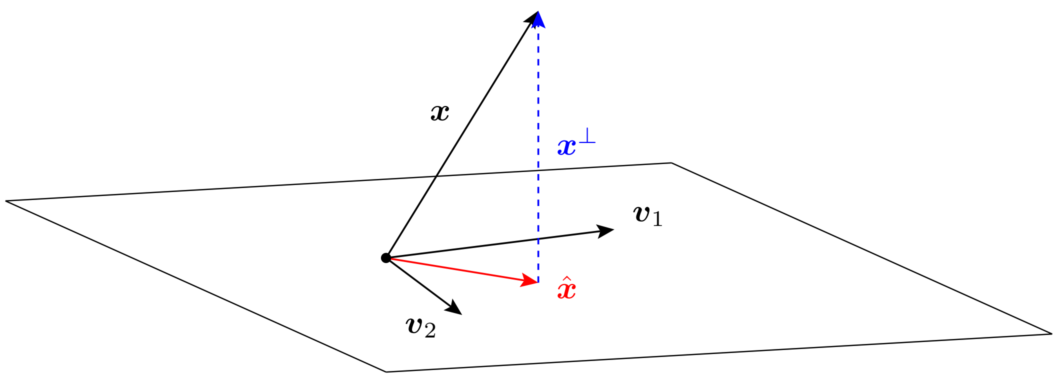
The experimental data might be of the form (v_i, w_i, y_i), where now the dependent variable y_i depends on two explanatory variables v_i, w_i (instead of just one x_i).

Example 3. Fitting a linear relationship $y_i \approx \beta_1 + \beta_2 v_i + \beta_3 w_i$, we get:

$$\underbrace{\begin{bmatrix} 1 & v_1 & w_1 \\ 1 & v_2 & w_2 \\ 1 & v_3 & w_3 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{design matrix}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\substack{\text{observation} \\ \text{vector}}}$$

And we again proceed by finding a least squares solution.

Review



- Suppose $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ is an orthonormal basis of W.
  The **orthogonal projection** of $\boldsymbol{x}$ onto W is:

$$\hat{\boldsymbol{x}} = \underbrace{\langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle \boldsymbol{v}_1}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_1} + \ldots + \underbrace{\langle \boldsymbol{x}, \boldsymbol{v}_m \rangle \boldsymbol{v}_m}_{\text{proj. of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_m}.$$

(To stay agile, we are writing $\langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle = \boldsymbol{x} \cdot \boldsymbol{v}_1$ for the inner product.)

Gram–Schmidt

Example 4. Find an orthonormal basis for $V = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Recipe. (Gram–Schmidt orthonormalization)

Given a basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, produce an orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$.

$$\begin{aligned} \boldsymbol{b}_1 &= \boldsymbol{a}_1, & \boldsymbol{q}_1 &= \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} \\ \boldsymbol{b}_2 &= \boldsymbol{a}_2 - \langle \boldsymbol{a}_2, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1, & \boldsymbol{q}_2 &= \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} \\ \boldsymbol{b}_3 &= \boldsymbol{a}_3 - \langle \boldsymbol{a}_3, \boldsymbol{q}_1 \rangle \boldsymbol{q}_1 - \langle \boldsymbol{a}_3, \boldsymbol{q}_2 \rangle \boldsymbol{q}_2, & \boldsymbol{q}_3 &= \frac{\boldsymbol{b}_3}{\|\boldsymbol{b}_3\|} \\ &\vdots \end{aligned}$$

Armin Straub
astraub@illinois.edu

Example 5. Find an orthonormal basis for $V = \operatorname{span}\left\{ \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \right\}$.

Solution.

$$\boldsymbol{b}_1 = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \qquad \boldsymbol{q}_1 = \begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}$$

$$\boldsymbol{b}_2 = \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix} - \left\langle \begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, \qquad \boldsymbol{q}_2 = \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}$$

$$\boldsymbol{b}_3 = \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} - \left\langle \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \boldsymbol{q}_1 \right\rangle \boldsymbol{q}_1 - \left\langle \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \boldsymbol{q}_2 \right\rangle \boldsymbol{q}_2 = \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}, \qquad \boldsymbol{q}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}$$

We have obtained an orthonormal basis for V:

$$\begin{bmatrix} 1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix}.$$