Today

- Finding the best fitting model y=ax.
- Using a spreadsheet.
- Finding the best fitting model y=ax+b.
- Other types of models.
- Note: WW Assignment 6 due Friday @5pm.

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Define f(a):

(A) $SSR(a) = |5-4a| + |7-6a|$

(B) $SSR(a) = (4-5a)^2 + (6-7a)^2$

(C) $SSR(a) = (5-4a)^2 + (7-6a)^2$

(D) $SSR(a) = (5-4-a)^2 + (7-6-a)^2$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Define f(a):

(A) $SSR(a) = |5-4a| + |7-6a|$

(B) $SSR(a) = (4-5a)^2 + (6-7a)^2$

(C) $SSR(a) = (5-4a)^2 + (7-6a)^2$

(D) $SSR(a) = (5-4-a)^2 + (7-6-a)^2$

Recall: $f(a) = (y_1-ax_1)^2 + (y_2-ax_2)^2$

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Find the a that minimizes SSR(a):

(A) a = 7/6

(B) a = 5/4

(C) a = (7/6 + 5/4) / 2

(D) a = 31/26

Find a so that y=ax fits (4,5), (6,7) in the "least squares" sense.

Find the a that minimizes SSR(a):

$$SSR(a) = (5-4a)^2 + (7-6a)^2$$
$$= 5^2 - 2 \cdot 4 \cdot 5a + 4^2a^2 + 7^2 - 2 \cdot 6 \cdot 7a + 6^2a^2$$
$$SSR'(a) = -2 \cdot 4 \cdot 5 + 2 \cdot 4^2 a - 2 \cdot 6 \cdot 7 + 2 \cdot 6^2 a = 0$$
$$a = (2 \cdot 4 \cdot 5 + 2 \cdot 6 \cdot 7) / (2 \cdot 4^2 + 2 \cdot 6^2)$$
$$= (4 \cdot 5 + 6 \cdot 7) / (4^2 + 6^2) = 62/52$$
$$= (x_1 \cdot y_1 + x_2 \cdot y_2) / (x_1^2 + x_2^2)$$

Notation

$$\sum_{i=1}^{n} q_i = q_1 + q_2 + \ldots + q_n$$

$$\sum_{i=1}^{n} (y_i - ax_i)^2 = (y_1 - ax_1)^2 + (y_2 - ax_2)^2 + \ldots + (y_n - ax_n)^2$$

Find a so that y=ax fits (x_1,y_1), (x_2,y_2),..., (x_n,y_n) in the "least squares" sense.

Find the a that minimizes SSR(a):

(A) $a = \sum_{i=1}^{n} y_i \,/\, \sum_{i=1}^{n} x_i$

(C) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i$

(B) $a = \sum_{i=1}^{n} x_i \,/\, \sum_{i=1}^{n} y_i$

(D) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i^2$

Find a so that $y=ax$ fits (x_1,y_1), (x_2,y_2),..., (x_n,y_n) in the "least squares" sense.

Find the a that minimizes SSR(a):

(A) $a = \sum_{i=1}^{n} y_i \,/\, \sum_{i=1}^{n} x_i$

(C) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i$

(B) $a = \sum_{i=1}^{n} x_i \,/\, \sum_{i=1}^{n} y_i$

(D) $a = \sum_{i=1}^{n} x_i y_i \,/\, \sum_{i=1}^{n} x_i^2$

Find a so that $y=ax$ fits (x_1,y_1), (x_2,y_2),…, (x_n,y_n) in the "least squares" sense.

$$SSR(a) = \sum_{i=1}^{n} (y_i - ax_i)^2$$

$$= \sum (y_i^2 - 2ax_iy_i + a^2x_i^2)$$

$$SSR'(a) = \sum (0 - 2x_iy_i + 2ax_i^2) = 0$$

$$-2x_1y_1 + 2ax_1^2 - 2x_2y_2 + 2ax_2^2 + \ldots = 0$$

$$ax_1^2 + ax_2^2 + \ldots = x_1y_1 + x_2y_2 + \ldots$$

$$a = (x_1y_1 + x_2y_2 + \ldots) / (x_1^2 + x_2^2 + \ldots)$$

Definitions

- A model is a function that you use to summarize or fit data. For example, some common ones: $f(x)=ax$, $f(x)=ax+b$, $f(x)=Ce^{-kx}$.
- Residuals are a measure of how far each model value is from the data value: $r_i=y_i-f(x_i)$.
- The Sum of Squared Residuals (SSR) is a measure of how well the model fits all the data: $SSR = \sum (y_i-f(x_i))^2$. Small is better.
- The best fit model is the model with parameter value(s) (a, a&b, C&k) that gives the smallest SSR.

For best fits using y=ax+b, see course notes supplement.

$$a = \frac{P_{avg} - \bar{x}\bar{y}}{X^2_{avg} - \bar{x}^2}$$

$$b = \bar{y} - a\bar{x}$$

$$P_{avg} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$X^2_{avg} = \frac{1}{n}\sum_{i=1}^{n} (x_i^2)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

You don't have to memorize this. It's just for reference. In fact, most spreadsheets have a function that does it for you.

- Using a spreadsheet...

- The rest of these slides were not covered in class but might help you get a better sense for when you might use various model for fitting data.

Examples of models for different types of data

Data: The rate of an enzyme's activity as a function of the concentration of enzyme.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

Examples of models for different types of data

Data: The rate of an enzyme's activity as a function of the concentration of enzyme.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

We now know how to do this.

Examples of models for different types of data

Data: The number of radioactive atoms left in a block of uranium after various times have elapsed.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

Examples of models for different types of data

Data: The number of radioactive atoms left in a block of uranium after various times have elapsed.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

Requires optimizing over two parameters (C and k) - can be done but not in MATH 102.

Examples of models for different types of data

Data: The number of calories you need to eat in a day depending on how long a run you take in the morning.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

Examples of models for different types of data

Data: The number of calories you need to eat in a day depending on how long a run you take in the morning.

A suitable model:

(A) y=m

(B) y=ax

(C) y=ax+b

(D) $y=Ce^{-kx}$

This also requires optimizing over two paramters but we have the formulae for this particular case.

Examples of models for different types of data

Data: The height of each student in the class.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

Examples of models for different types of data

Data: The height of each student in the class.

A suitable model:

(A) y=m

(B) y=ax

(C) y=ax+b

(D) $y=Ce^{-kx}$

$$SSR(m) = (h_1-m)^2 + (h_2-m)^2 + \ldots + (h_n-m)^2$$

$$= h_1^2 - 2h_1m + m^2 + h_2^2 - 2h_2m + m^2 + \ldots + h_n^2 - 2h_nm + m^2$$

$$SSR'(m) = -2h_1 + 2m - 2h_2 + 2m + \ldots - 2h_n + 2m = 0$$

Examples of models for different types of data

Data: The height of each student in the class.

A suitable model:

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

$$SSR(m) = (h_1-m)^2 + (h_2-m)^2 + \dots + (h_n-m)^2$$

$$= h_1^2 - 2h_1m + m^2 \dots$$

$$SSR'(m) = -h_1 + m \dots$$

The “best fit” constant model is the mean of the data!

$$nm = h_1 + h_2 + \dots + h_n$$

$$m = (h_1 + h_2 + \dots + h_n) / n$$

Examples of models for different types of data

(A) $y=m$

(B) $y=ax$

(C) $y=ax+b$

(D) $y=Ce^{-kx}$

All of these models can be "best fit" by minimizing the SSR, called least squares analysis.

When you use $y=ax+b$, it's got a special name as well: linear regression.

If the minimization problem is quadratic in the parameters (e.g. A,B,C), it's called linear least squares. Otherwise (e.g. D), it's called nonlinear least squares.