Least squares: intro to fitting a line to data

Math 102 Section 102
Mingfeng Qiu

Oct. 17, 2018

Announcements

- Solution for the worksheet last week has been uploaded.
  - If you find a mistake, let me know.
- Keep up with your required coursework and extra practice.
- I'll do whatever I can to help you.
- I'm not here on Friday except for the lecture, neither can I answer emails.
  - Help for OSH 4: Dr. Lisanne Rens @MATX 1118, 2-3 pm

Today...

1. Time concepts
2. Finish up Optimal Foraging
3. Measuring central tendency: average
4. Fitting a line to data

Time

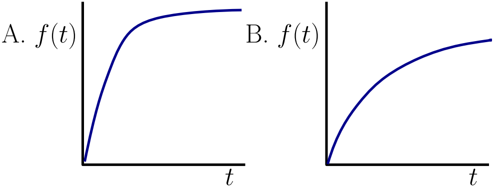
Two kinds of time that we talk about:

- A period of time: cannot be negative
- A moment: relative to a certain event. Can be negative

Bear eating berries

Q6. $t = \sqrt{k\tau}$.

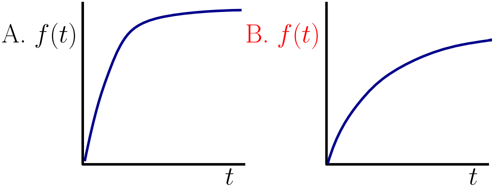
There are two different patches which take the same amount of time τ to get to. In which patch should the bear spend more time in order to maximize the average energy gain?



$f(t) = \frac{t}{20+t}$ $f(t) = \frac{t}{200+t}$

Bear eating berries

Q6. Recall that $t = \sqrt{k\tau}$.
There are two different patches which take the same amount of time τ to get to. In which patch should the bear spend more time in order to maximize the average energy gain?



$f(t) = \frac{t}{20+t}$ $\qquad$ $f(t) = \frac{t}{200+t}$

If τ is fixed, but there are two patches, one with k_1 and one with k_2. The bear must stay in patch with the bigger k_i longer, to optimize the average rate of energy gain.

Average energy gain

Recall:

- Average energy gain per unit time:

$$R(t) = \frac{\text{energy gained}}{\text{total time spent}}$$

- energy gained $= f(t)$
- total time spent $=$ travel time $+$ time at patch $= \tau + t$

$$R(t) = \frac{f(t)}{t + \tau}.$$

Optimizing the average energy gain

$$R'(t) = \frac{f'(t)(t+\tau) - f(t)}{(\tau+t)^2}$$

Q7. The critical points of $R(t)$ satisfy

A. $t = \sqrt{k\tau}$

B. $t_{1,2} = \pm\sqrt{k\tau}$

C. $f'(t) = \frac{f(t)}{\tau+t}$

D. $f'(t) = 0$, $f(t) = 0$

E. $f(t)(\tau+t) = f'(t)$

Optimizing the average energy gain

The optimal time t satisfies

$$R'(t) = \frac{f'(t)(\tau + t) - f(t)}{(\tau + t)^2} = 0,$$

which is the same as

$$f'(t) = \frac{f(t)}{\tau + t} = R(t)$$

The optimal time t is the time at which the instantaneous rate of food collection ($f'(t)$) equals the average rate of food collection $R(t)$.

Marginal Value Theorem

Marginal Value Theorem: To maximize energy intake, the optimal time to stay at a patch of food with energy collection function $f(t)$ with travel time τ is

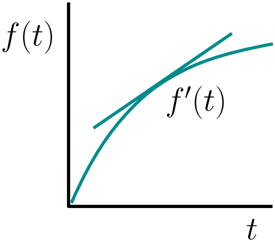
$$\underbrace{f'(t)}_{\text{instantaneous rate}} = \underbrace{R(t)}_{\text{average rate}}$$

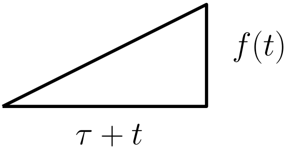
- The rate of benefit of a given resource is maximized by exploiting the resource until the rate of benefit falls to the maximum average rate that can be sustained over a long period.

Marginal Value Theorem: Geometry

$$f'(t) = \frac{f(t)}{\tau + t} = R(t)$$

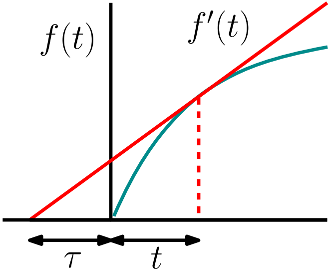
- $f'(t)$ = slope of tangent line
- $\frac{f(t)}{\tau+t}$ = ratio





Marginal Value Theorem: Geometry

$$f'(t) = \frac{f(t)}{\tau + t}$$



Summary for optimal foraging

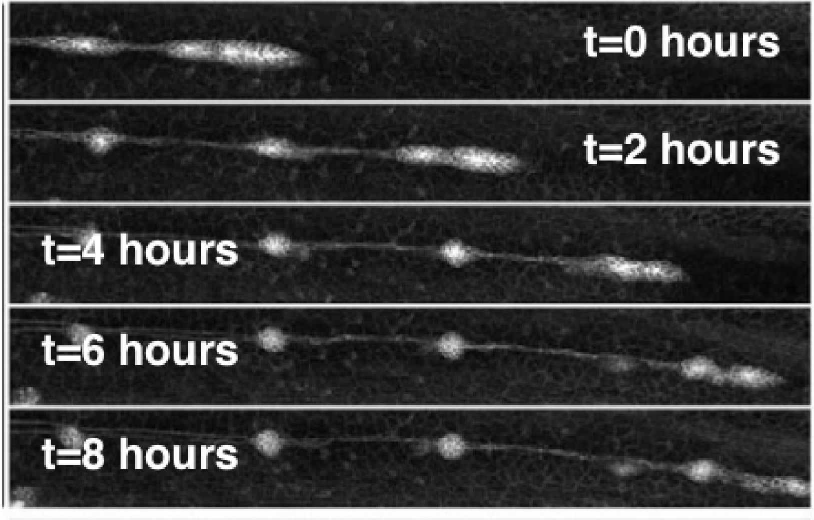
On Monday and today we

- developed a simple model for how an animal collects food (gains energy) in a patch.
- looked at the various types of food patches, with different $f(t)$.
- calculated the optimal time for a specific example (Bear eating berries)
- discovered the Marginal Value Theorem and interpreted it geometrically

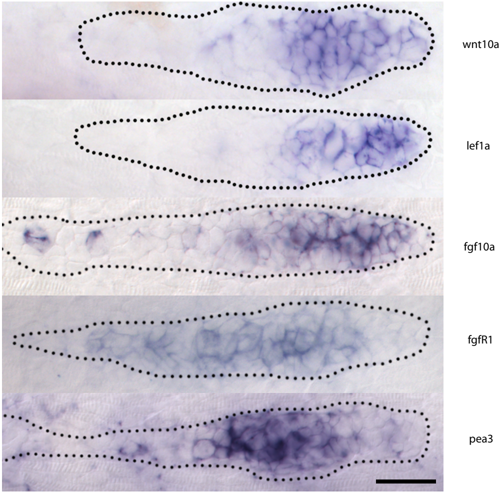
Fitting data: least squares

The content about least squares is found on Canvas > Main Resources > Calendar (weekly schedule) > Week 7 supplement.
Link here

Zebrafish posterior lateral line primordium (PLLP)

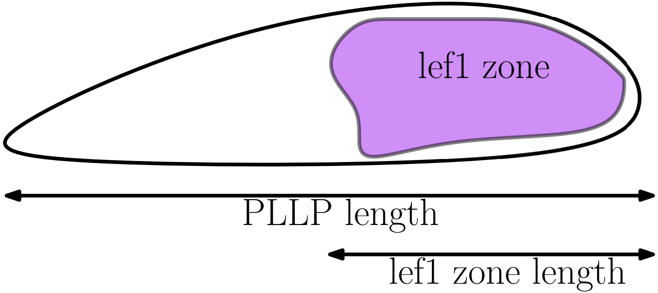


Zebrafish PLLP



Ajay Chitnis, Damian Dalle Nogare, NIH

Size of lef1 domain relative to the PLLP



Size of lef1 domain relative to the PLLP

- Let
$$x = \frac{\text{length of lef1 zone}}{\text{length of PLLP}}.$$
- In four experiments, the ratio of the lef1 domain to the PLLP was found to be
$$x_1 = 0.84, x_2 = 0.49, x_3 = 0.72 \text{ and } x_4 = 0.53.$$

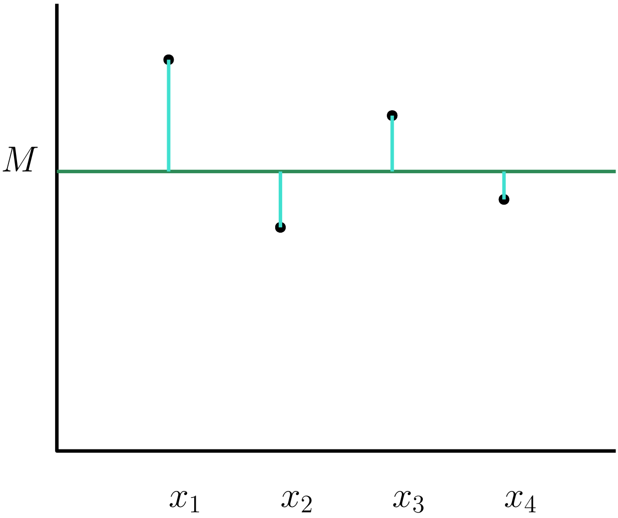
Measuring central tendency

- Data set: x_1, x_2, x_3, x_4
- Goal: Find a number M that summarizes the data set
- Idea: Find M such that the sum of squared residuals (SSR) is as small as possible

$$f(M) = \sum_{i=1}^{4}(x_i - M)^2 = (x_1 - M)^2 + (x_2 - M)^2 + (x_3 - M)^2 + (x_4 - M)^2$$

- Each $r_i = x_i - M$ is called a residual.

Measuring central tendency



Measuring central tendency

- Minimize $f(M)$
- Critical Points

$$
\begin{aligned}
f'(M) &= -2(x_1 - M) - 2(x_2 - M) - 2(x_3 - M) - 2(x_4 - M) \\
&= -2(x_1 + x_2 + x_3 + x_4) + 8M \\
M &= \frac{x_1 + x_2 + x_3 + x_4}{4} \quad \text{when } f'(M) = 0
\end{aligned}
$$

- This turns out to be the global minimum.
- It is the average of the data points!

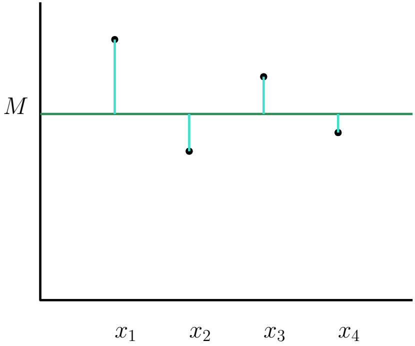
Average size

The average relative size of lef1 domain is

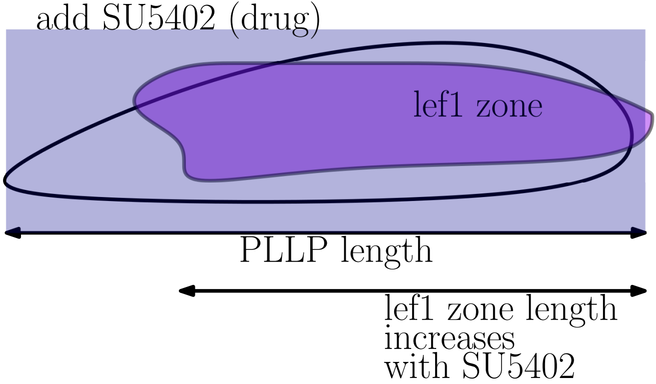
$$\begin{aligned} M &= \frac{x_1 + x_2 + x_3 + x_4}{4} \\ &= \frac{0.89 + 0.49 + 0.72 + 0.53}{4} \\ &= 0.6575. \end{aligned}$$

Measuring central tendency

- The average, or mean, results from minimizing the SSR for horizontal line to data:
$$M = \frac{\sum_{i=1}^{n} x_i}{n}$$



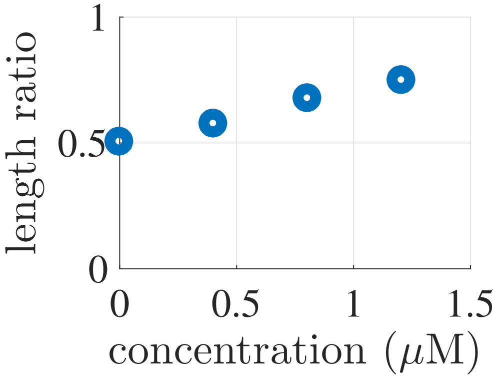
Size of lef1 domain relative to the PLLP



- The size of the lef1 domain increases with the concentration of SU5402 added

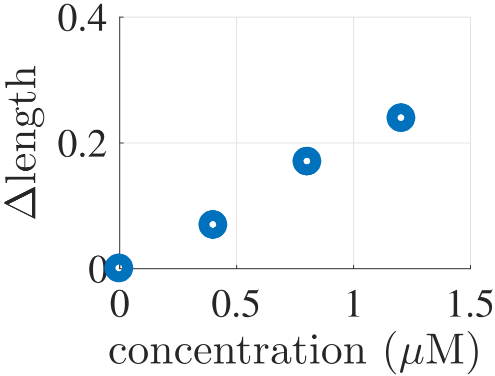
Fitting a line to data

SU5402 concentration μM	0	0.4	0.8	1.2
lef1/PLLP length ratio	0.51	0.58	0.68	0.75



Fitting a line without intercept

SU5402 concentration μM	0	0.4	0.8	1.2
change in length ratio	0	0.07	0.17	0.24



Fitting a line without intercept

Fact (Line fitting without intercept)

Suppose we have n data points (x_i, y_i) where $i = 1, 2, \cdots, n$, which are fit by a line through the origin $y = ax$. The SSR is

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

The value of a that minimizes the SSR is

$$a = \frac{\sum_{i=1}^{n}(x_i y_i)}{\sum_{i=1}^{n} x_i^2}.$$

Fitting a line without intercept to data

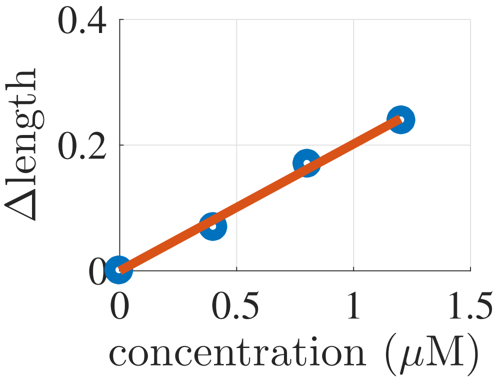
x_i	0.4	0.8	1.2
y_i	0.07	0.17	0.24
r_i	$0.07 - 0.4a$	$0.17 - 0.8a$	$0.24 - 1.2a$

$$\begin{aligned} a &= \frac{\sum_{i=1}^{n}(x_i y_i)}{\sum_{i=1}^{n} x_i^2} \\ &= \frac{x_1 y_1 + x_2 y_2 + x_3 y_3}{x_1^2 + x_2^2 + x_3^2} \\ &= \frac{0.4 \cdot 0.07 + 0.8 \cdot 0.17 + 1.2 \cdot 0.24}{0.4^2 + 0.8^2 + 1.2^2} \\ &= \frac{0.452}{2.24} = \frac{113}{560} \approx 0.2. \end{aligned}$$

- Best fit line:

$$y = 0.2x$$

Fitting a line without intercept to data



$$y = 0.2x$$

Answers

6. B
7. C