

Báo cáo cuối kỳ vấn đáp

- Code:
 - Một ứng dụng dashboard cập nhật dữ liệu mới tối đa mỗi phút.
 - Có thể sửa thông số biểu đồ/đồ thị một cách có giới hạn và phù hợp.
 - Khuyến khích dùng biểu đồ tương tác.
 - Dùng ít nhất 02 bảng dữ liệu (có kết hợp bảng).
 - Desktop app/web app/Zeppelin notebook.
- Báo cáo file word (20 – 40 trang):
 - Trình bày dữ liệu, nội dung dashboard, C.nghệ sử dụng,...
 - Ít nhất 02 cơ sở lập luận (có biểu đồ/đồ thị) và 02 kết luận.

Tuần 4: Biểu đồ và đồ thị (tiếp theo)

Giảng viên: ThS. Lê Minh Tân

Nội dung

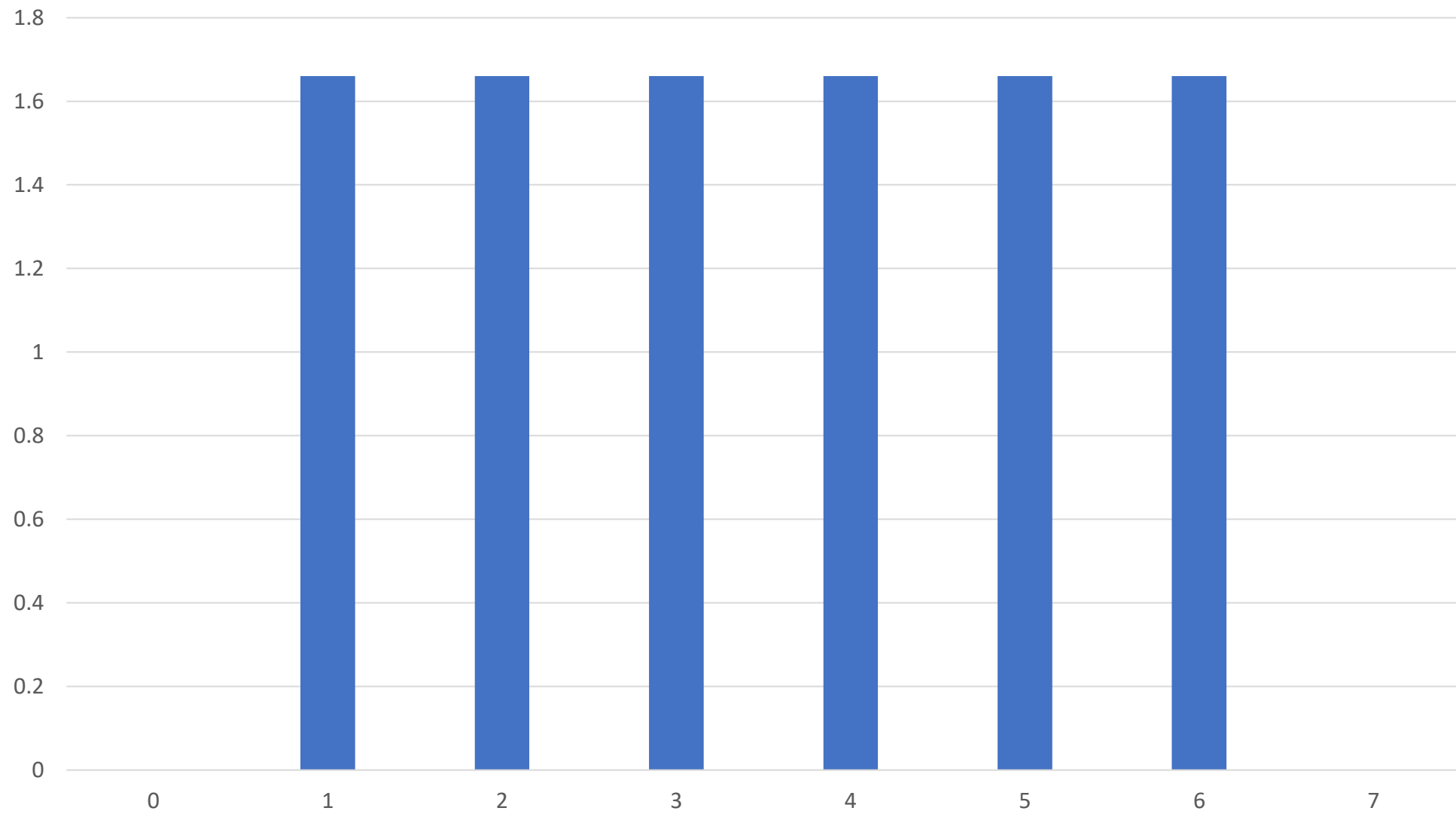
- I. Phân phối và histogram
- II. Vẽ biểu đồ bằng Python
- III. Vẽ đồ thị
- IV. Zeppelin nâng cao

I. Phân phối và histogram

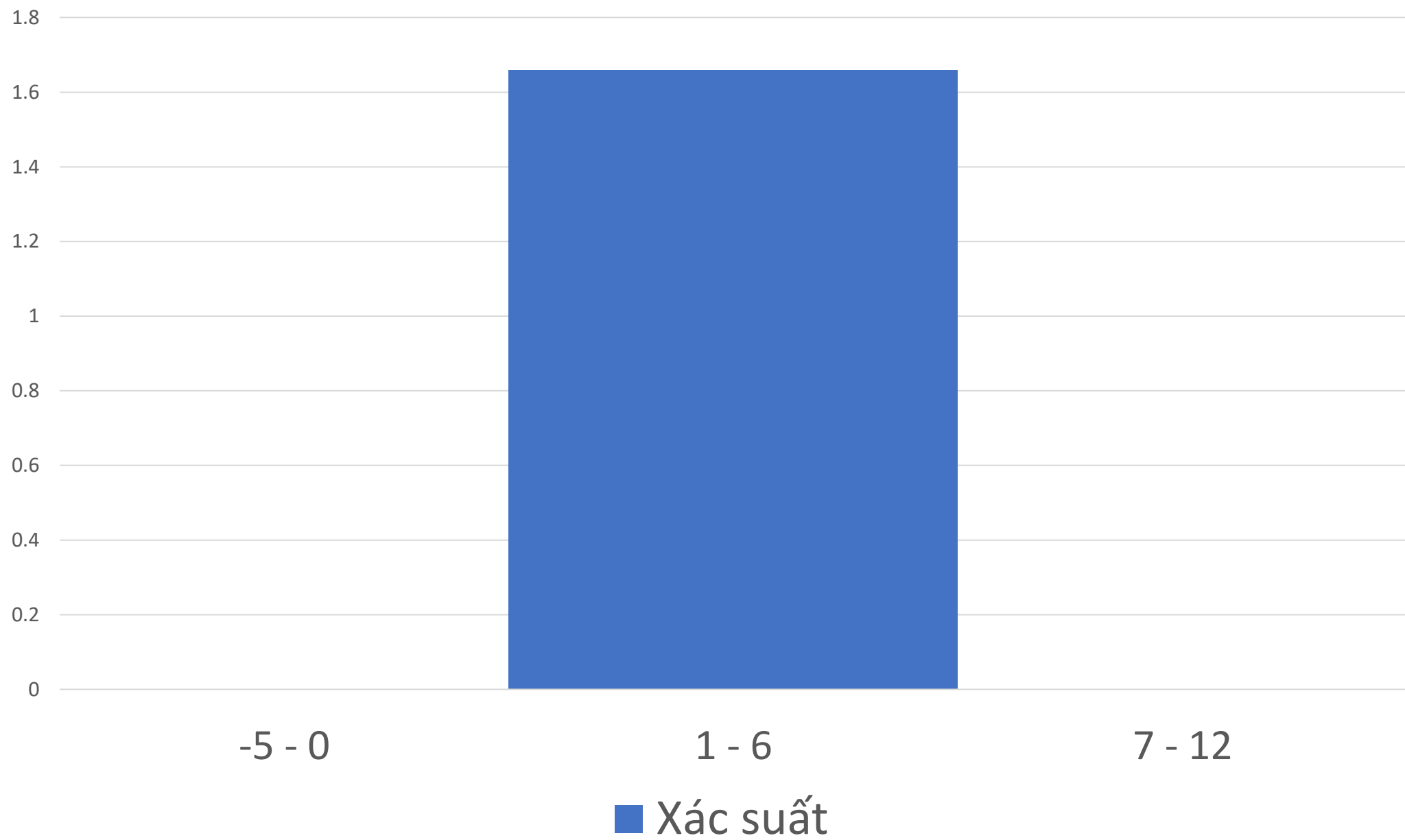
- Phân phối thể hiện tần suất sự kiện của biến.
- Phân phối có thể được biểu diễn bằng một hàm $f(x)$.

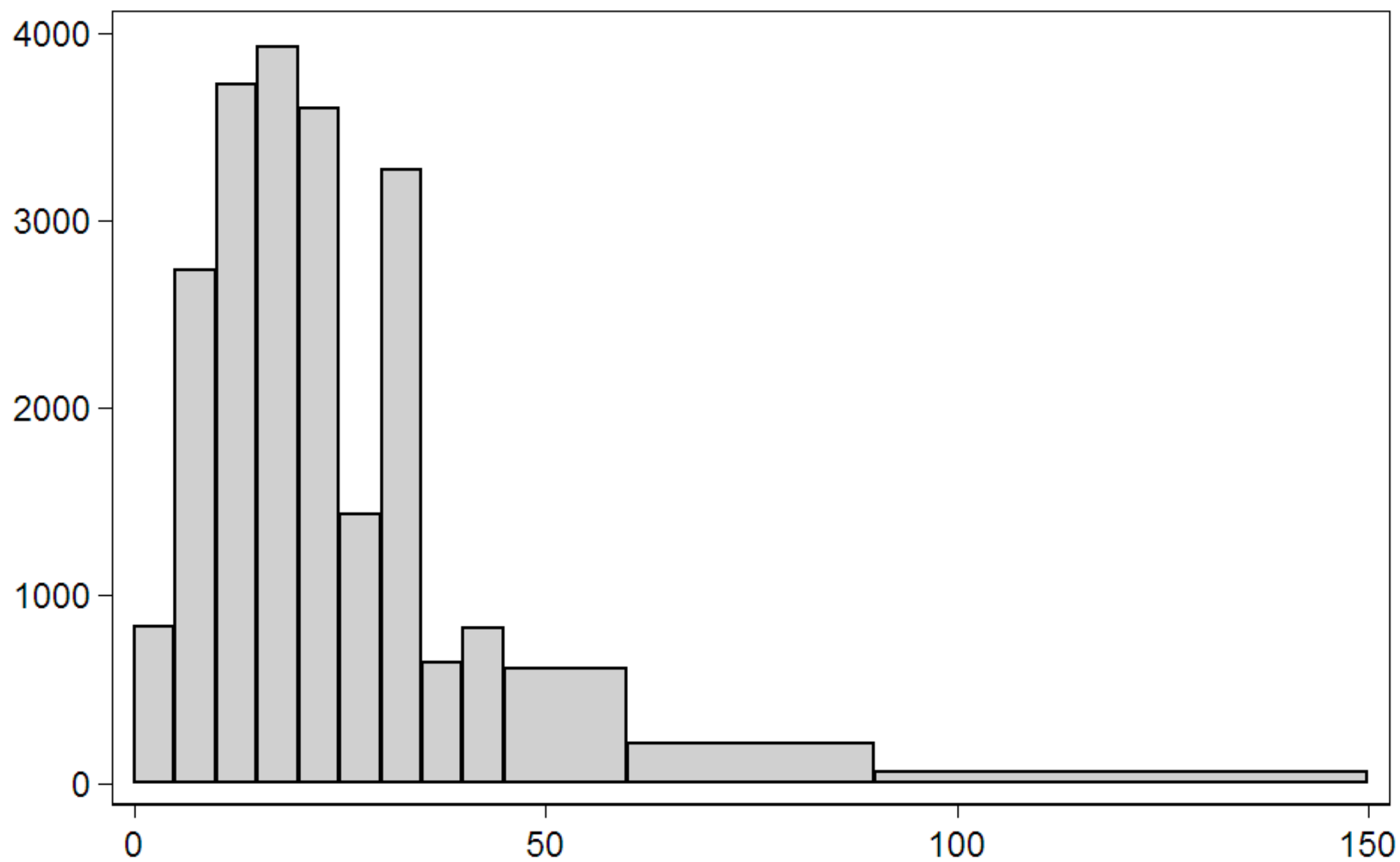
$$f(x) = \begin{cases} \frac{1}{6} & \text{if } 1 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

| x | f(x) |
|-----|------|
| ... | 0 |
| 1 | 1/6 |
| 2 | 1/6 |
| 3 | 1/6 |
| 4 | 1/6 |
| 5 | 1/6 |
| 6 | 1/6 |
| ... | 0 |



■ Xác suất





Xây dựng histogram:

1. Định nghĩa xô (lớp) dữ liệu
 1. Số lượng xô c ?
 2. Khoảng dữ liệu r ?
 3. Độ rộng cột của phân lớp w ?
2. Vẽ biểu đồ
 1. Tính tổng phần tử theo xô

$$\min_x$$

$$r = \max_x - \min_x$$

$$w = \frac{r}{c}$$

$$y_i = |X_i|$$

Gợi ý số lượng xô c :

$$c = \sqrt{n}$$

Quy tắc Sturge:

$$c = 1 + 3.322 \log_{10} n$$

Dùng quy tắc Sturge khi:

- Phân phối đều
- Dữ liệu phân phối cân đối quanh giá trị trung bình

II. Vẽ biểu đồ bằng Python

- **matplotlib, seaborn**
 - Biểu đồ tĩnh
- **plotly**
 - Biểu đồ tương tác

| | Matplotlib | Plotly | Seaborn |
|---|----------------------------|--------------------------------------|--|
| Chức năng | Vẽ biểu đồ/đồ thị | Vẽ biểu đồ/đồ thị | <ul style="list-style-type: none"> - Trực quan hóa dữ liệu - Hỗ trợ matplotlib |
| Ngôn ngữ | Python | Python, R, JS,... | Python |
| Hàm vẽ | plt.<type> plt.subplots | go.Figure (Python) plot_ly (R) | sns.relplot |
| Tương thích Zeppelin 0.8.2 | 3.5.2 Python >= 3.10 | Mới nhất | Mới nhất |

III. Vẽ đồ thị

- Bước 1: Định nghĩa hàm
- Bước 2: Tạo các điểm dữ liệu từ hàm bước 1
- Bước 3: Vẽ đồ thị

IV. Zeppelin nâng cao

- Revision
- Trình thông dịch html
- Job và cron