

§1. NHỮNG KHÁI NIỆM MỞ ĐẦU

§2. TRÌNH BÀY MỘT MẪU SỐ LIỆU

A. TÓM TẮT LÝ THUYẾT.

1. Khái niệm về thống kê

Thống kê là khoa học về các phương pháp thu thập, tổ chức, trình bày, phân tích và xử lý số liệu.

2. Mẫu số liệu

- **Dấu hiệu** là một vấn đề hay hiện tượng nào đó mà người điều tra quan tâm tìm hiểu. Mỗi đối tượng điều tra gọi là một **đơn vị điều tra**. Mỗi đơn vị điều tra có một số liệu, số liệu đó gọi là giá trị của dấu hiệu trên đơn vị điều tra đó.

- Một tập con hữu hạn các đơn vị điều tra được gọi là một **mẫu**. Số phần tử của một mẫu được gọi là **kích thước mẫu**. Các giá trị của dấu hiệu thu được trên mẫu được gọi là một **mẫu số liệu** (mỗi giá trị như thế còn gọi là một số liệu của mẫu).

- Nếu thực hiện điều tra trên mọi đơn vị điều tra thì đó là điều tra toàn bộ. Nếu chỉ điều tra trên một mẫu thì đó là điều tra mẫu.

3. Bảng phân bố tần số - tần suất. Bảng phân bố tần số - tần suất ghép lớp.

Tần số của giá trị x_i là số lần lặp lại của giá trị x_i trong mẫu số liệu.

Tần suất f_i của giá trị x_i là tỷ số giữa tần số n_i và kích thước mẫu N hay $f_i = \frac{n_i}{N}$.

Người ta thường viết tần suất dưới dạng phần trăm.

- Bảng phân bố tần số (gọi tắt là **bảng tần số**) được trình bày ngang như sau:

Giá trị (x)	x_1	x_2	x_3	...	x_m	
Tần số (n)	n_1	n_2	n_3	...	n_m	$N = \sum_{i=1}^m n_i$

Trên hàng tần số, người ta dành một ô để ghi kích thước mẫu N bằng tổng các tần số (tức $N = \sum_{i=1}^m n_i$).

- Bổ sung thêm một hàng tần suất vào bảng trên, ta được **bảng phân bố tần số - tần suất** (gọi tắt là **bảng tần số - tần suất**).

Giá trị (x)	x_1	x_2	x_3	\dots	x_m	
Tần số (n)	n_1	n_2	n_3	\dots	x_m	$N = \sum_{i=1}^m n_i$
Tần suất %	f_1	f_2	f_3	\dots	f_m	

Chú ý: Người ta cũng thể hiện bảng phân bố tần số - tần suất dưới dạng bảng dọc.

• Nếu kích thước mẫu số liệu khá lớn, thì người ta thường chia số liệu thành nhiều lớp dưới dạng $[a;b]$ hay $[a;b)$ (thường có độ dài các lớp bằng nhau). Khi đó tần số của lớp $[a;b]$ là số giá trị $x_i \in [a;b]$ (hay $x_i \in [a;b)$) xuất hiện trong lớp đó. Tần suất của lớp $[a;b]$ là $f = \frac{n}{N}$ trong đó n là tần số của lớp $[a;b]$ và N là kích thước mẫu.

- Bảng phân bố tần suất ghép lớp được xác định tương tự như trên.

- Giá trị đại diện của lớp $[a;b]$ là $c = \frac{a+b}{2}$

4. Biểu đồ:

Các loại biểu đồ thường dùng là: biểu đồ hình cột, biểu đồ đường gấp khúc và biểu đồ hình quạt. Số liệu vẽ biểu đồ được lấy từ các bảng tần số - tần suất.