



TIN HỌC ỨNG DỤNG

(CH2 - CÁC HÀM THỐNG KÊ)

Phan Trọng Tiến

BM Công nghệ phần mềm

Khoa Công nghệ thông tin, VNUA

Email: phantien84@gmail.com

Website: <http://timoday.edu.vn>



Nội dung chính

1. Khái niệm công thức
2. Khái niệm về hàm
3. Cách nhập công thức và hàm
4. Một số hàm thống kê cơ bản
5. Chi tiết một số hàm thống kê



1. Công thức (nhắc lại)

- ❑ Công thức
 - ❑ Bắt đầu bởi dấu “=“
 - ❑ Sau đó:
 - ❑ Địa chỉ, hằng, miền,...
 - ❑ Toán tử
 - ❑ Hàm
- ❑ VD:
 - ❑ = A1+A2-B2
 - ❑ = SIN(A1) + COS(B2)
 - ❑ = LN(A5)



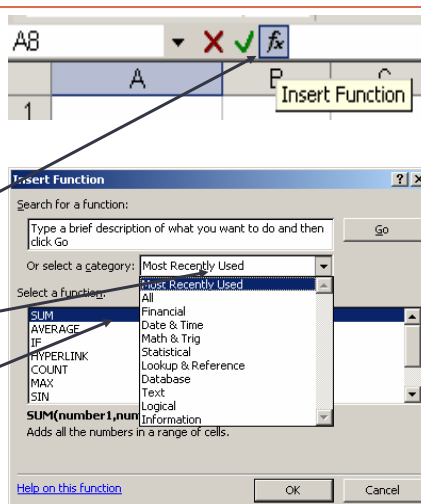
2. Khái niệm hàm

- ❑ Là các hàm tính toán được xây dựng sẵn.
- ❑ Phục vụ các tính toán thông dụng.
- ❑ Cú pháp:
 - Tên_hàm** (danh_sách_đối_số)
- ❑ Đối số được phân cách bởi dấu phẩy
 - ❑ Ví dụ: =**Average**(miền_dữ_liệu)
- ❑ Đối số có thể là giá trị, địa chỉ, hằng,...



3. Nhập công thức và hàm

- ❑ Nhập trực tiếp vào ô
- ❑ Sử dụng thanh công thức
 - ❑ Kích chuột vào biểu tượng f_x để mở hộp thoại chọn hàm.
- ❑ Select a category: loại hàm.
- ❑ Select a function: chọn hàm.
- ❑ Hoặc từ menu **Formulas**/Insert Function



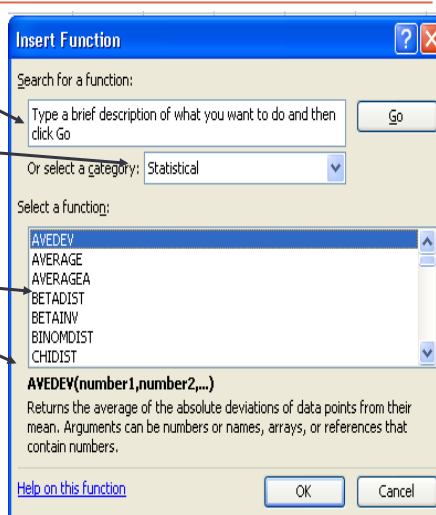
Ch2 - Các hàm thống kê

5



Chọn các hàm thống kê

- ❑ Đánh vào đây để tìm kiếm hàm
- ❑ Chọn theo phân loại hàm
- ❑ Các hàm theo phân loại
- ❑ Giải thích về hàm



Ch2 - Các hàm thống kê

6



4. Một số hàm thống kê cơ bản

- ❑ AVERAGE → Cho trung bình của miền
- ❑ CHIINV → Cho trị của hàm χ về phân bố xác suất
- ❑ CHITEST → Cho kết quả kiểm tra χ về tính độc lập của 2 dãy số
- ❑ CORREL → Cho hệ số tương quan giữa 2 dãy số
- ❑ COUNT → Đếm số ô có giá trị là số trong miền
- ❑ COUNTA → Đếm số ô có dữ liệu
- ❑ COVAR → Cho hiệp phương sai (covariance)
- ❑ FINV → Cho hệ số Fisher
- ❑ FREQUENCY → Cho tần số phân bố của cột số

Ch2 - Các hàm thống kê

7



4. Một số hàm thống kê cơ bản

- ❑ FTEST → Cho kết quả kiểm tra Fisher (F-test)
- ❑ LINEST → Cho các tham số A, B của hàm hồi quy tuyến tính $Y = Ax + B$
- ❑ LOGNEST → Cho các tham số của hàm hồi quy dạng mũ $Y = (B * (M1^{X1}) * (M2^{X2}) * \dots)$
- ❑ MAX → Cho trị cực đại của miền
- ❑ MIN → Cho trị cực tiểu của miền
- ❑ STDEV → Cho độ lệch chuẩn
- ❑ TINV → Cho hệ số Student
- ❑ TTEST → Cho kết quả kiểm tra Student (T-test)
- ❑ VAR → Cho phương sai của miền

Ch2 - Các hàm thống kê

8



5. Chi tiết một số hàm thống kê

1. Các hàm AVERAGE, COUNT, MAX, MIN, STDEV, VAR
2. Hàm CORREL và COVAR
3. Hàm FINV
4. Hàm FREQUENCY
5. Hàm LINEST
6. Hàm TINV



5.1 Các hàm AVERAGE, COUNT, MAX, MIN, STDEV, VAR

- ❑ Cú pháp: **Tên_Hàm**(Miền_DL)
 - ❑ **AVERAGE** : trung bình cộng của miền
 - ❑ **COUNT** : đếm số lượng các ô dl là số
 - ❑ **MAX** : cho cực đại của miền
 - ❑ **MIN** : cho cực tiểu của miền
 - ❑ **STDEV** : độ lệch chuẩn là một đại lượng thống kê mô tả dùng để đo mức độ phân tán của một tập dữ liệu
 - ❑ **VAR** : phương sai là đo sự phân tán thống kê của biến đó, nó hàm ý các giá trị của biến đó thường ở cách giá trị kỳ vọng bao xa.
 - ❑ Ví dụ: =VAR(C1:C50)



5.2 Hàm CORREL và COVAR

- ❑ Cú pháp: Tên_Hàm(**dãy_số1, dãy_số2**)
 - ❑ CORREL: hệ số tương quan cho biết độ mạnh của mối tương quan tuyến tính giữa hai biến số ngẫu nhiên
 - ❑ COVAR: **hiệp phương sai** là độ đo sự biến thiên cùng nhau của hai biến ngẫu nhiên



Chú ý

- ❑ Dãy số phải là kiểu số
- ❑ Nếu giá trị trong dãy số khác kiểu số thì giá trị này bỏ qua.
- ❑ Nếu hai dãy số khác nhau về kích thước thì gặp lỗi #N/A
- ❑ Nếu một trong hai dãy rỗng thì gặp lỗi #DIV/0



Ví dụ

	A	B	C
1	Data1	Data2	
2	3	9	=CORREL(A2:A6,B2:B6)
3	2	7	=COVAR(A2:A6,B2:B6)
4	4	12	
5	5	15	
6	6	17	

Ch2 - Các hàm thống kê

13



5.3 Hàm FINV

- Cho nghịch đảo của phân bố xác suất theo quy luật Fisher. Có thể dùng phân bố F trong một F-test để so sánh độ phân tán của hai nhóm số liệu.
- Cú pháp
FINV(mức_xác_xuất, bậc_tự_do1, bậc_tự_do_2)

↙
Là xác suất gắn với phân
bố Fisher thường = 0.05
hay 0.01

↘
Là bậc tự do tử số

↘
Là bậc tự do mẫu số

Ch2 - Các hàm thống kê

14



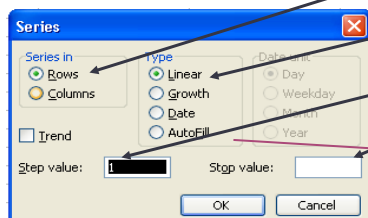
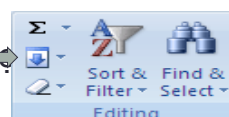
5.4 Hàm ***FREQUENCY***

- ❑ Tính tần xuất các giá trị trong một khoảng dữ liệu và ***trả về một mảng các số theo chiều dọc***
- ❑ Miền phân tổ là miền định khoảng phân chia dữ liệu
- ❑ Ví dụ: khoảng 1-> 100, được chia thành 10 khoảng con vậy có cách nào làm tự động?



Cách tạo miền phân tổ tự động

- ❑ Nhập giá trị đầu 1 vào ô đầu của miền.
- ❑ Vào Home/Edit/Fill/Series...
một cửa sổ hiện ra cho phép bạn lựa chọn



Tạo theo hàng hay theo cột

Kiểu dữ liệu miền phân tổ

Bước tăng

Giá trị dừng

Cấp số cộng
Cấp số nhân
Ngày tháng
Tự động điền



Cách sử dụng

- ❑ =FREQUENCY(miền_d_liệu,miền_p_tổ)
- ❑ Chú ý
 - ❑ Kết quả trả về một mảng công thức sau khi bạn chọn một miền liên kế các ô.
 - ❑ Số phần tử trả về lớn hơn một phần tử so với số phần tử miền phân tổ.
 - ❑ Hàm bỏ qua các ô trống hoặc văn bản
 - ❑ Các công thức được nhập như một mảng, cách làm: Đánh công thức, rồi chọn miền có cả công thức, ấn **F2** rồi ấn **Ctrl+Shift+Enter**

Ch2 - Các hàm thống kê

17



Ví dụ tính tần xuất điểm

	A	B
1	Scores	Bins
2	79	70
3	85	79
4	78	89
5	85	
6	50	
7	81	
8	95	
9	88	
10	97	

Scores<=70
 70<Scores<=79
 79<Scores<=89
 Scores >89

Formula	Description (Result)
=FREQUENCY(A2:A10,B2:B5)	Number of scores less than or equal to 70 (1)
	Number of scores in the bin 71-79 (2)
	Number of scores in the bin 80-89 (4)
	Number of scores greater than or equal to 90 (2)

Ch2 - Các hàm thống kê

18



5.5 Hàm **LINEST**

- ❑ Tính thống kê cho đường hồi quy tuyến tính. Kết quả trả về **một mảng** miêu tả đường thẳng này.
- ❑ Phương trình đường thẳng
 - ❑ $y = m \cdot x + b$
 - ❑ hoặc $y = m_1 \cdot x_1 + m_2 \cdot x_2 + \dots + b$
 - ❑ Giá trị phụ thuộc y là một hàm của các giá trị độc lập x , m_i là các hệ số tương ứng với giá trị x_i , b là giá trị hằng số.
- ❑ LINEST cũng có thể trả về thêm các thống kê hồi quy.



Cú pháp

- ❑ **LINEST**(*known_y's*,*known_x's*,*const*,*stats*)
 - ❑ **known_y's** là tập hợp các giá trị đã biết trong quan hệ đường $y = mx + b$.
 - ❑ **known_x's** là một tùy chọn, tập hợp các giá trị đã biết trong mối quan hệ đường $y = mx + b$.
 - ❑ Nếu **known_x's** bỏ qua, thì nó lấy mảng mặc định $\{1,2,3,\dots\}$ bằng kích thước **known_y's**.
 - ❑ **const** là giá trị logic xác định tác động lên giá trị b
 - ❑ **True** hoặc **bỏ qua** thì b được **tính bình thường**.
 - ❑ **False** thì $b = 0$, đường thẳng lúc đó chỉ là $y = mx$
 - ❑ **stats** là giá trị logic xác định có trả về thêm thống kê hồi quy hay không
 - ❑ **True**: trả về thêm các thống kê hồi quy, mảng được trả về $\{mn, mn-1, \dots, m1, b; sen, sen-1, \dots, se1, seb; r2, sey; F, df; ssreg, ssresid\}$.
 - ❑ **False** hoặc **bỏ qua**: chỉ trả về hệ số m và hằng số b



Ví dụ

	A	B	C	D	E
1	Floor space (x1)	Offices (x2)	Entrances (x3)	Age (x4)	Assessed value (y)
2	2310	2	2	20	142,000
3	2333	2	2	12	144,000
4	2356	3	1.5	33	151,000
5	2379	3	2	43	150,000
6	2402	2	3	53	139,000
7	2425	4	2	23	169,000
8	2448	2	1.5	99	126,000
9	2471	2	2	34	142,900
10	2494	3	3	23	163,000
11	2517	4	4	55	169,000
12	2540	2	3	22	149,000
	Formula				
	=LINEST(E2:E12,A2:D12,TRUE,TRUE)				

Ch2 - Các hàm thống kê

21



Kết quả

- ❑ Trả về giá trị -234.24
- ❑ Công thức phải được nhập như một mảng, cách nhập công thức như một mảng: chọn vùng A14:E18 (chọn cả ô có công thức)
- ❑ Ấn **F2** và ấn **CTRL+SHIFT+ENTER**
- ❑ $y = 27.64 \cdot x_1 + 12,530 \cdot x_2 + 2,553 \cdot x_3 - 234.24 \cdot x_4 + 52,318$

-234.237	2553.211	12529.77	27.64139	52317.83
----------	----------	----------	----------	----------

Ch2 - Các hàm thống kê

22



5.6 Hàm **TINV**

- ❑ Cho nghịch đảo của phân bố xác suất theo Student với số bậc tự do nhất định.
- ❑ Cú pháp: **TINV(mức_xác_xuất,số_bậc_tự_do)**
 - ❑ Mức_xác_xuất: là xác suất gắn với phân bố Student
 - ❑ Số_bậc_tự_do: là số bậc tự do cần để đặc trưng phân bố
 - ❑ $TINV = p(t < X)$, với X là biến ngẫu nhiên
- ❑ Ví dụ $TINV(0.05,12) = 2.18$