XỬ LÝ SỐ LIỆU TRONG EXCEL

Giới thiệu về Data Analysis

Trong Bảng tính Excel có một phần chuyên xử lý số liệu, gọi là Data Analysis, tuy chưa thật sâu nhưng trong tình hình hiện tại có thể coi là đủ dùng để xử lý thống kê các số liệu thu thập được trong quá trình điều tra nghiên cứu (Tính các đặc trưng thống kê cơ bản, chia tổ, vẽ biểu đồ, tính hệ số tương quan, hiệp phương sai, tính và vẽ đường hồi quy tuyến tính hoặc phi tuyến, tính hồi quy bội tuyến tính, vẽ đồ thị kiểu cột, kiểu bánh tròn, làm trơn số liệu v.v...) và trong các kiểu bố trí thí nghiệm với một hoặc hai nhân tố (Phân tích phương sai một nhân tố, hai nhân tố, so sánh 2 phương sai, so sánh hai trung bình v v ...). Cách vào số liệu và chọn công cụ xử lý đơn giản, dễ dùng, đồ hoạ đẹp. Tuy nhiên vì quan niệm người dùng đã biết cách xử lý số liệu và hiểu được các kết quả nên Excel chỉ in ra các kết quả tóm tắt, không giải thích gì thêm. Phần Help có tỷ mỷ hơn nhưng cũng rất vắn tắt.

Trước khi dùng phải có số liệu, tuỳ vấn đề mà sắp xếp số liệu cho thích hợp, sau đó vào Menu Tools chọn Data Analysis (Nếu không thấythì phải mở Add -ins sau đó bổ sung thêm Analysis Toolpak, nếu không thấy Analysis ToolPak thì phải Setup lại Excel để bổ sung).

Data Analysis	×
<u>A</u> nalysis Tools	ок (
Anova: Single Factor	
Anova: Two-Factor With Replication	Cancel
Anova: Two-Factor Without Replication	
Correlation	
Covariance	
Descriptive Statistics	
Exponential Smoothing	
F-Test Two-Sample for Variances	
Fourier Analysis	
Histogram	<u>•</u>

Menu Data Analysis có dạng như sau:

Chọn trong Menu công cụ xử lý thích hợp ta được một hộp thoại, nhìn chung mỗi hộp thoại gồm 3 phần: Phần Input Range để ghi dịa chỉ miền số liệu cần xử lý, phần Options để có các chọn lựa thích hợp, cuối cùng là phần Output Range để chọn nơi in ra kết quả.

Thí dụ Chọn Anova Single Factor (Phân tích phương sai một nhân tố) được hộp thoại Input Range: ghi địa chỉ miền vào

Options : Số liệu để theo cột thì đánh dấu Columns, để theo hàng thì chọn Rows, có

tên các mức thì chọn Labels, chọn mức ý nghĩa Alpha

Output Range : ghi địa chỉ một vùng trắng trong Sheet để ghi kết quả

hoặc chọn 1 trang mới (New Worksheet)

hoặc một tệp mới (New Work Book)

Anova: Single Factor		×
Input Input Range: Grouped By: Labels in First Row Alpha: 0.05	Columns C <u>R</u> ows	OK Cancel <u>H</u> elp
Output options © <u>O</u> utput Range: © New Worksheet <u>Ply</u> : © New <u>W</u> orkbook		

Có thể chia Data Analysis ra thành 5 nhóm:

1/ Thống kê mô tả: Thống kê mô tả: **Descriptive** Statistics Nhât đồ : Histogram Trung bình trượt: Moving Average Làm trơn số liêu: **Exponential Smoothing** Thứ hạng và phân vị: Rank and percentile 2/ So sánh: So sánh hai phương sai của 2 mẫu quan sát: F- test two sample for means So sánh hai trung bình khi lấy mẫu theo cặp: T- test Paired two samples for means So sánh hai trung bình khi lấy mẫu độc lập T- test two sample assuming giả thiết phương sai bằng nhau: equal variances So sánh hai trung bình khi lấy mẫu độc lập T- test two sample assuming giả thiết phương sai khác nhau: unequal variances So sánh hai trung bình khi biết phương sai: Z- test two sample for means 3/ Phân tích phương sai: Phân tích phương sai một nhân tố Anova single factor Phân tích phương sai hai nhân tố không lặp lai Anova two factor without replication Phân tích phương sai hai nhân tố có lặp lai Anova two factor with replication 4/ Hiệp phương sai, tương quan, hồi quy: Hiệp phương sai *Covariance* Correlation Tương quan Hồi quy Regression 5/ Một số tiện ích Lấy mẫu Sampling Phân tích Fourrier Fourrier Analysis

Random number generation

Tạo số ngẫu nhiên

Bài1 THỐNG KÊ MÔ TẢ

I/ Nhật đồ_(Histogram)

Khi có nhiều số liệu, để trong một cột hay để trong một bảng, chúng ta muốn chia khoảng, tính các tần số ứng với mỗi khoảng sau đó vẽ nhật đồ để xem số liệu có phân phối chuẩn không thì dùng Histogram.

Các bước làm như sau:

- 1- Để số liệu trong 1 cột hay bảng chữ nhật
- 2- Tìm giá trị lớn nhất bằng (hàm Max), giá trị nhỏ nhất (hàm Min)
 - $l\hat{a}y = Max Min$
- 3- Chọn số khoảng n (trong thực tế thường chọn n từ 20 30)
- 4- Tìm h = R / n (Để bớt lẻ có thể dùng hàm Round§ (h, số số lẻ)
- 5- Tạo cột Bin sau đó gọi Histogram

Thí dụ chiều dài 30 con cá

Max = 49 Min = 11 R = 38 n = 8 h = 38/5 quy tròn h = 5

Tạo cột Bin Xuất phát gần Min thí dụ 12 tiếp theo lấy 12 + 5 = 17 $17 + 5 = 22 \dots$ Cho đến sát Max (49)

	A	В	С	D	E	F	G	H	Ι	J	K	L	M
1		chiêu d	ài 30 co	n cá	bin		Bin	equen	icy .	ľ			
2	14	21	27		12		12	1			н	istogram	
3	25	27	34		16		16	2					
4	17	16	37		20		20	5		6 . 5 T	_		
5	20	11	35		24		24	3					
6	35	19	40		28		- 28	б		9 3-	3		∎Frequ
1	38	22	26		32		32	4		Ĩ <u>ĕ</u> 2†			
8	40	49	26		36		36	4		" 1	, I I,II,II,II,		Д
9	25	30	34		40		40	4		2°	່ 2	8 4	່ 🧧
10	32	24	19		44		44	0					Š.
11	31	29	20		48		48	0			В	in	
12							More	1		ļ		-	

Histogram		×
⊡nput Input Range: Bin Range: IT Labels	\$A\$2:\$C\$11 \$E\$2:\$E\$11	OK Cancel <u>H</u> elp
Output options Output Range: Output Range: New Worksheet <u>Ply:</u> New <u>W</u> orkbook Pareto (sorted histogram) Cumulative Percentage Ohart Output	J1	

Nếu muốn nhật đồ sắp xếp theo thứ tự tăng dần thì chọn Pareto, nếu muốn vẽ thêm đường tần suất luỹ tích (cộng dồn) thì chọn cumulative percentage.

2/ Thống kê mô tả (Descriptive Statistics)

Khi có một bảng gồm nhiều cột, mỗi cột là một biến, tên biến đặt ở dòng đầu thì có thể tính ngay tất cả các thống kê cho tất cả các biến bằng cách gọi thống kê mô tả (cũng có thể sắp xếp các số liệu theo hàng, mỗi hàng là một biến).

Các công việc cần làm:

Chọn miền vào (Bảng gồm nhiều cột, mỗi cột là một biến, các biến không nhất thiết phải dài bằng nhau, nhưng khi khai báo Input Range thì phải khai hình chữ nhật bao trùm toàn bộ bảng).

Khai báo số liệu theo cột hay hàng.

Chọn nhãn (nháy vào ô Label) nếu tên biến đặt ở dòng đầu.

Chọn số to thứ mấy (k-Largest), nếu chọn 1 (mặc định) thì có số to nhất (Max), nếu chọn 2 thì có số to thứ nhì v v...

Chọn số nhỏ thứ mấy (k- Smallest), nếu chọn 1 (mặc định) thì có số nhỏ nhất (Min), nếu chọn 2 thì có số nhỏ thứ nhì v v . . .

Sau đó khai báo miền ra, quan trọng nhất là phải chọn ô **Summary Statistics** để có được các thống kê.

Descriptive Statistics		?>
_Input		
Input Range:	\$B\$46:\$E\$61 💽	ОК
Grouped By:	Columns	Cancel
Grouped by:	O Rows	Help
Labels in First Row	C Tour	
Output options		1
• Output Range:	\$B\$64 🗾	
O New Worksheet Ply:		
C New <u>W</u> orkbook		
Summary statistics		
Confidence Level for Me	ean: 95 %	
Kth L <u>a</u> rgest:	1	
	1	

Thí dụ ta có bảng gồm 4 biến số đặt ở A2..D17

Sau khi khai báo cho in kết quả ra F2

Mỗi cột biến bây giở sẽ ứng với 2 cột kết quả, cột đầu ghi tên các thống kê, cột sau ghi giá trị của các thống kê, vì các cột tên giống nhau nên có thể để lại 1 cột còn xoá bớt, sau đó ghép các cột giá trị lại sát nhau cho đẹp.

X1	X2	X3	Y		X1	X2	X3	Y
52	40	81	5.5	Mean	45.33333	63.4	69.53333	10.93333
33	37	90	2.1	Standard Error	5.317238	6.137162	3.526588	1.69657
72	95	66	20.5	Median	46	62	70	10.3
15	58	40	9.6	Mode	#N/A	#N/A	#N/A	#N/A
40	20	75	1.7	Standard Deviation	20.59357	23.76913	13.65842	6.570787
32	41	80	3.8	Sample Variance	424.0952	564.9714	186.5524	43.17524
76	54	83	10.3	Kurtosis	-0.99416	-0.23096	0.616884	-0.23649
10	85	70	11.7	Skewness	-0.15496	0.116913	-0.81968	0.440507
68	70	65	15.2	Range	66	89	50	22.7
57	109	45	24.4	Minimum	10	20	40	1.7
24	62	64	9.3	Maximum	76	109	90	24.4
46	75	71	13	Sum	680	951	1043	164
35	55	82	6.5	Count	15	15	15	15
54	68	63	13.8	Confidence	11.40435	13.16292	7.563786	3.638784
				Level(95.0%)				
66	82	68	16.6					

Mean - Giá trị trung bình (còn tên khác là Average)Standard error - Sai số của trung bìnhMedian - Trung vịSample variance - phương sai mẫuKurtosis - Độ nhọnSkewness- Độ nghiêng (bất đối xứngb)

Range - Biên độ Maximum - Giá trị lớn nhất Skewness- Độ nghiêng (bất đối xứngb) Minimum - Giá trị nhỏ nhất Sum- Tổng (N/A là không xác định được) Count - Số quan sát

Thí dụ chung

Mở tệp Baitap1.xls. Chọn biến Dobeo

Confidence Interval - Khoảng tin cậy

Data Analysis		X
<u>A</u> nalysis Tools		ОК
Anova: Single Factor Anova: Two-Factor With Replication Anova: Two-Factor Without Replication Correlation Covariance Descriptive Statistics Exponential Smoothing F-Test Two-Sample for Variances Fourier Analysis Histogram		Cancel

Histogram			×
Input Input Range: Bin Range: Labels	\$A\$1:\$A\$121	%	OK Cancel <u>H</u> elp
Output options Output Range: New Worksheet Ply: New Workbook Pareto (sorted histogram) Cumulative Percentage Chart Output	\$H\$1		

