



TIN HỌC ỨNG DỤNG

(CH4 - PHÂN TÍCH PHƯƠNG SAI, SO SÁNH VÀ KIỂM ĐỊNH)

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. Phân tích phương sai
2. Kiểm định sự bằng nhau của hai mẫu
3. So sánh trung bình



1. Phân tích phương sai

- ❑ Dùng để phân tích các số liệu khi theo dõi ảnh hưởng của nhân tố và ảnh hưởng tương tác của chúng.
- ❑ Phân tích một nhân tố: bố trí thí nghiệm theo ô vuông La tinh.
- ❑ Phân tích hai nhân tố: bố trí thí nghiệm theo khối ngẫu nhiên, kiểu trực giao, kiểu chia ô lớn, ô vừa, ô nhỏ hoặc kết hợp vừa chia bằng vừa chia ô.
- ❑ Từ ba nhân tố trở lên: bố trí thí nghiệm sao cho mỗi nhân tố có hai mức hay mỗi nhân tố có ba mức.



1.1 Phân tích phương sai một nhân tố

- ❑ Dùng để phân tích số liệu khi theo dõi ảnh hưởng của các mức nhân tố tới kết quả như của các công thức cho ăn tới năng xuất thịt ...
- ❑ Thiết kế thí nghiệm kiểu hoàn toàn ngẫu nhiên, mỗi mức lặp lại một số lần, số lần lặp lại của các mức không cần phải bằng nhau.
- ❑ Số liệu đưa vào theo hàng hoặc theo cột (theo hàng thì mỗi hàng ứng với một mức nhân tố), ô đầu ghi tên mức, ô tiếp ghi SL.



Kiểm định và bố trí dữ liệu

Ta cần kiểm định giả thuyết về ảnh hưởng giống nhau của các tác nhân:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_t$$

$$H_1 : \exists h \neq k \text{ với } h, k = 1, 2, \dots, t \text{ mà } \mu_h \neq \mu_k$$

Dữ liệu của Thiết kế ngẫu nhiên hoàn toàn có dạng:

Mẫu của tác nhân				
1	2	3	...	t
X_{11}	X_{21}	X_{31}		X_{t1}
X_{12}	X_{22}	X_{32}		X_{t2}
.	.	.	.	.
.	.	.	..	.
X_{1n_1}	X_{2n_2}	X_{3n_3}		X_{tn_t}

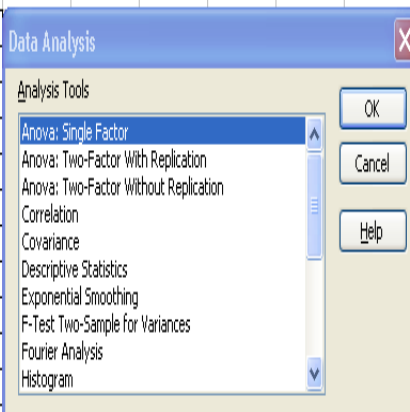
Ch4 - Phân tích phương sai, so sánh và kiểm định

5



Phân tích ảnh hưởng của các loại thuốc đến năng suất lúa.

	A	B	C	D	E	F	G	H	I	J	K
1	Ảnh hưởng của các loại thuốc đến năng suất lúa										
2	Loại thuốc	NS ô 1	NS ô 2	NS ô 3	NS ô 4						
3	T1	3.187	4.61	3.562	3.217						
4	T2	3.39	2.875	2.775							
5	T3	2.797	3.001	2.505	3.49						
6	T4	2.832	3.103	3.448	2.255						
7	T5	2.233	2.743	2.727							
8	T6	2.952	2.272	2.47							
9	T7	2.858	2.895	2.458	1.723						
10	T8	2.308	2.335	1.957							
11	T9	2.013	1.788	2.248	2.115						
12	T10	3.202	3.06	2.24	2.69						
13	T11	1.192	1.652	1.075	1.03						
14											



Ch4 - Phân tích phương sai, so sánh và kiểm định

6

Vào **Tools > Data Analysis > Anova: Single Factor**

Miền dữ liệu

Nhóm dữ liệu theo cột hay theo hàng

Đặt nhãn đầu dòng

Mức ý nghĩa

Nơi đặt kết quả

Ch4 - Phân tích phương sai, so sánh và kiểm định 7

Kết quả

Anova: Single Factor

SUMMARY					
Groups	Count	Sum	Average	Variance	
T1	4	14.576	3.644	0.443686	
T2	3	9.04	3.013333333	0.108908333	
T3	4	11.793	2.94825	0.17187425	
T4	4	11.638	2.9095	0.253933667	
T5	3	7.703	2.567666667	0.084065333	
T6	3	7.694	2.564666667	0.122321333	
T7	4	9.934	2.4835	0.296197667	
T8	3	6.6	2.2	0.044469	
T9	4	8.164	2.041	0.037706	
T10	4	11.192	2.798	0.184962667	
T11	4	4.949	1.23725	0.08111425	

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	15.10396	10	1.510395928	8.541711422	2.66E-06	2.176844
Within Groups	5.127952	29	0.176825914			
Total	20.23191	39				

Ch4 - Phân tích phương sai, so sánh và kiểm định 8



Phân tích kết quả

- ❑ Kết quả in ra gồm các thống kê cơ bản cho từng công thức (trung bình, độ lệch chuẩn ...)
- ❑ Nếu giá trị xác suất **P-value < alpha** (hay **F_{tn} > F_{lt}**) → công thức có tác động khác nhau tới kết quả, ngược lại các công thức không có khác biệt đáng kể.
- ❑ Nếu các công thức có tác động khác nhau → tiếp tục so sánh các công thức có giống nhau hay khác nhau hay không.



Phân tích kết quả

- ❑ Ví dụ: $F = 8.541 > F_{crit} = 2.176$ → các công thức có tác động khác nhau tới năng suất lúa. Muốn so sánh xem công thức nào có ảnh hưởng khác nhau tới TB không -> dùng phương pháp so sánh dùng chỉ số LSD.
- ❑ Nhận xét: Công thức T1 cho năng suất cao nhất. Tiếp tục so sánh các công thức có giống nhau hay khác nhau hay không ta dùng so sánh trung bình bằng chỉ số LSD.



1.2 So sánh trung bình dùng chỉ số LSD (Least Significance Difference)

- ❑ Excel không cho phép so sánh các trung bình của các nhóm ứng với các mức của nhân tố (công thức)
- ❑ Nếu cần so sánh trung bình CT T_i (với r_i lần lặp) với trung bình CT T_j (với r_j lần lặp) có thể tính thêm chỉ số

$$LSD = t_{\alpha, f} * SQRT(s^2(1/r_i + 1/r_j))$$

- ❑ s^2 là phương sai chung ước lượng bởi trung bình sai số bình phương trong nội bộ nhóm (*MS within groups*)
- ❑ $\alpha = 1 - p$
- ❑ $t_{\alpha, f}$ là giá trị t của bảng Student (Hàm *TINV*)



So sánh

- ❑ Tính trị tuyệt đối của các trung bình T_i, T_j : $|T_i - T_j|$
- ❑ So sánh nếu $|T_i - T_j| > LSD$ thì hai trung bình là khác nhau, ngược lại hai trung bình được coi là không khác nhau.
- ❑ Thường người ta lập bảng hiệu các trung bình, sau đó lập bảng so sánh.



Ví dụ so sánh ảnh hưởng của thuốc T_1, T_2 đến năng xuất lúa

Groups	Count	Sum	Average	$T_1 - T_i$
T1	4	14.576	3.644	
T2	3	9.04	3.013333333	0.630666667
T3	4	11.793	2.94825	0.69575
T4	4	11.638	2.9095	0.7345
T5	3	7.703	2.567666667	1.076333333
T6	3	7.694	2.564666667	1.079333333
T7	4	9.934	2.4835	1.1605
T8	3	6.6	2.2	1.444
T9	4	8.164	2.041	1.603
T10	4	11.192	2.798	0.846
T11	4	4.949	1.23725	2.40675

Ch4 - Phân tích phương sai, so sánh và kiểm định

13



Ví dụ

- ❑ $|T_1 - T_2| = |3.6440 - 3.0133| = 0.6307$
- ❑ $T(0.05, 29) = T_{inv}(0.05, 29) = 2.045$
- ❑ $S^2 = 0.17682$
- ❑ LSD được tính trong TH r_i, r_j bằng 3 hoặc bằng 4

$$LSD = 2.045 * \text{SQRT}(0.17682 * (1/3 + 1/4)) = 0.656739049$$

$$LSD = 2.045 * \text{SQRT}(0.17682 * (1/4 + 1/4)) = 0.608022212$$

$$LSD = 2.045 * \text{SQRT}(0.17682 * (1/3 + 1/3)) = 0.702083575$$

Trường hợp T_1, T_2 :

$|T_1 - T_2| = 0.6307 < LSD = 0.656739049$ nên CThức T_1, T_2 không khác nhau rõ rệt. Tương tự các TH khác

T_1 cho năng suất cao nhất là tốt nhất, công thức T_{11} cho năng suất thấp nhất là kém nhất.

Ch4 - Phân tích phương sai, so sánh và kiểm định

14



1.3 Phân tích phương sai hai nhân tố

- ❑ Xảy ra hai trường hợp:
 - ❑ Nhân tố A và B không tương tác, biến động gây nên bởi tác động đồng thời của A và B gần sát 0.
 - ❑ Nhân tố A và B có tương tác.



a. Phân tích phương sai hai nhân tố không tương tác

- ❑ Bố trí số liệu: hàng là các mức của nhân tố thứ nhất, cột là các mức của nhân tố thứ hai.

Dữ liệu của thiết kế khối ngẫu nhiên có dạng:

Các tác nhân chính (Treatments)	Các tác nhân ngoại lai (Blocks)			
	B ₁	B ₂	...	B _b
T ₁	$\bar{x}_{11}$	$\bar{x}_{12}$	...	$\bar{x}_{1b}$
T ₂	$\bar{x}_{21}$	$\bar{x}_{22}$	...	$\bar{x}_{2b}$
...	...	...	...	...
T _t	$\bar{x}_{t1}$	$\bar{x}_{t2}$	...	$\bar{x}_{tb}$

Ta cần kiểm định giả thuyết về ảnh hưởng giống nhau của các tác nhân chính:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_t$$

$$H_1 : \exists h \neq k \text{ với } h, k = 1, 2, \dots, t \text{ mà } \mu_h \neq \mu_k$$



Ví dụ

- Ví dụ: nhân tố 1 có 4 mức, nhân tố 2 có 4 khối

	A	B	C	D	E
1		Khối 1	Khối 2	Khối 3	Khối 4
2	Mức 1	47	52	62	51
3	Mức 2	50	54	67	57
4	Mức 3	57	53	69	57
5	Mức 4	54	65	74	59

Ch4 - Phân tích phương sai, so sánh và kiểm định

17



Vào *Data>Data Analysis>Anova: Two Factor Without Replication*

	A	B	C	D	E	F	G
1		Khối 1	Khối 2	Khối 3	Khối 4		
2	Mức 1	47	52	62	51		
3	Mức 2	50	54	67	57		
4	Mức 3	57	53	69	57		
5	Mức 4	54	65	74	59		
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							

Anova: Two-Factor Without Replication

Input

Input Range:

☒ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

Ch4 - Phân tích phương sai, so sánh và kiểm định

18

Kết quả

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
M?c 1	4	212	53	40.66667
M?c 2	4	228	57	52.66667
M?c 3	4	236	59	48
M?c 4	4	252	63	74
Kh?i 1	4	208	52	19.33333
Kh?i 2	4	224	56	36.66667
Kh?i 3	4	272	68	24.66667
Kh?i 4	4	224	56	12

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	208	3	69.33333	8.914286	0.004648	3.862548
Columns	576	3	192	24.68571	0.000112	3.862548
Error	70	9	7.777778			
Total	854	15				

Ch4 - Phân tích phương sai, so sánh và kiểm định

19

Ch4 - Phân tích phương sai, so sánh và kiểm định



Phân tích kết quả

- ❑ Các mức của nhân tố 1 có ảnh hưởng khác nhau đến kết quả ($F_{tn} > F_{lt}$)
- ❑ Các mức của nhân tố 2 có ảnh hưởng khác nhau đến kết quả ($F_{tn} > F_{lt}$)
- ❑ Có thể tính F_{lt} bằng hàm FINV
 - ❑ = FINV(0.05,3,9)
- ❑ Có thể tính T bằng hàm TINV
 - ❑ = TINV(0.05,9)

20

Ch4 - Phân tích phương sai, so sánh và kiểm định



b. Phân tích phương sai hai nhân tố tương tác

- ❑ Công cụ hữu ích với dữ liệu đã được phân loại theo không gian hai chiều.
- ❑ Ví dụ: Thí nghiệm đo chiều cao của cây dừa, bằng cách dùng các công thức bón phân khác nhau (A,B,C) và nhiệt độ khác nhau (cao,thấp) -> 6 cặp {phân bón, nhiệt độ} chúng ta có một số quan sát chiều cao của cây.



Bố trí dữ liệu

Dữ liệu của thiết kế thừa số ngẫu nhiên hai yếu tố chính có dạng:

<i>Các tác nhân của yếu tố thứ nhất (A)</i>	<i>Các tác nhân của yếu tố thứ hai (B)</i>			
	B_1	B_2	...	B_b
A_1	$x_{11(1)}$ $x_{11(2)}$... $x_{11(c)}$	$x_{12(1)}$ $x_{12(2)}$... $x_{12(c)}$	...	$x_{1b(1)}$ $x_{1b(2)}$... $x_{1b(c)}$
...	...	...	...	...
A_a	$x_{a1(1)}$ $x_{a1(2)}$... $x_{a1(c)}$	$x_{a2(1)}$ $x_{a2(2)}$... $x_{a2(c)}$	...	$x_{ab(1)}$ $x_{ab(2)}$... $x_{ab(c)}$



Cần kiểm định

Giả thuyết (A) về ảnh hưởng giống nhau của các tác nhân thuộc nhân tố A:

$$H_0 : \mu_{A1} = \mu_{A2} = \dots = \mu_{At}$$

$$H_1 : \exists h \neq k \text{ với } h, k = 1, 2, \dots, t \text{ mà } \mu_{Ah} \neq \mu_{Ak}$$

Giả thuyết (B) về ảnh hưởng giống nhau của các tác nhân thuộc nhân tố B:

$$H_0 : \mu_{B1} = \mu_{B2} = \dots = \mu_{Bb}$$

$$H_1 : \exists h \neq k \text{ với } h, k = 1, 2, \dots, t \text{ mà } \mu_{Bh} \neq \mu_{Bk}$$

Giả thuyết (C) không có tương tác giữa các tác nhân của hai nhân tố:

$$H_0 : \text{Không có tương tác giữa các tác nhân của hai yếu tố}$$

$$H_1 : \text{Có sự tương tác giữa các tác nhân của hai yếu tố}$$



Ví dụ ns lúa của nhân tố phân bón (A) và mật độ trồng (B)

4		15	17	19
5		16	19	18
6		21	18	17
7	C thức 2	20	18	21
8		19	19	20
9		23	18	21
10		19	20	23
11	C thức 3	21	21	21
12		19	22	22
13		22	21	18
14		20	23	21

 **Vào *Data*Data Analysis**

	B	C	D	E	F	G	H	I
	Năng xuất bông (tạ/ha)							
	Mức 1	Mức 2	Mức 3					
	14	15	19					
	15	17	19					
	16	19	18					
	21	18	17					
	20	18	21					
	19	19	20					
	23	18	21					
	19	20	23					
	21	21	21					
	19	22	22					
	22	21	18					
	20	23	21					
	20	21	24					
	23	22	23					
	21	19	21					
	19	20	25					

Data Analysis

Analysis Tools

- 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

Ch4 - Phân tích phương sai, so sánh và kiểm định **25**

 **Hộp thoại *Anova: Two Factor With Replication***

Miền dữ liệu

Số hàng trên một mẫu

Chọn nơi để kết quả

Anova: Two-Factor With Replication

Input

Input Range: \$A\$2:\$D\$18

Rows per sample: 4

Alpha: 0.05

Output options

☒ Output Range: \$H\$5

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

Ch4 - Phân tích phương sai, so sánh và kiểm định **26**



Kết quả

Anova: Two-Factor With Replication

SUMMARY	M?c 1	M?c 2	M?c 3	Total
<i>C thức 1</i>				
Count	4	4	4	12
Sum	66	69	73	208
Average	16.5	17.25	18.25	17.33333
Variance	9.666667	2.916667	0.916667	4.242424
<i>C thức 2</i>				
Count	4	4	4	12
Sum	81	75	85	241
Average	20.25	18.75	21.25	20.08333
Variance	3.583333	0.916667	1.583333	2.810606
<i>C thức 3</i>				
Count	4	4	4	12
Sum	82	87	82	251
Average	20.5	21.75	20.5	20.91667
Variance	1.666667	0.916667	3	1.901515
<i>C thức 4</i>				
Count	4	4	4	12
Sum	83	82	93	258
Average	20.75	20.5	23.25	21.5
Variance	2.916667	1.666667	2.916667	3.727273

Ch4 - Phân tích phương sai, so sánh và kiểm định 27



Kết quả (tiếp)

Total						
Count	16	16	16			
Sum	312	313	333			
Average	19.5	19.5625	20.8125			
Variance	6.8	4.395833	5.095833			
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Sample	122.4167	3	40.80556	14.9898	1.71E-06	2.866266
Columns	17.54167	2	8.770833	3.221939	0.051615	3.259446
Interaction	23.95833	6	3.993056	1.466837	0.21725	2.363751
Within	98	36	2.722222			
Total	261.9167	47				

Ch4 - Phân tích phương sai, so sánh và kiểm định 28



Phân tích kết quả

- ❑ $F_A = 14.9898 > F_{0.5A} = 2.86627 \rightarrow$ nhân tố phân bón ảnh hưởng rõ rệt tới năng suất bông.
- ❑ $F_B < F_{0.5B} \rightarrow$ tăng mật độ cây không hy vọng tăng năng suất bông.
- ❑ $F_{AB} < F_{0.5AB} \rightarrow$ ảnh hưởng đồng thời của cả hai nhân tố không tác động đáng kể tới năng suất bông.



2. Kiểm định sự bằng nhau của hai mẫu

- ❑ Dùng để so sánh hai mẫu thông qua kiểm định giả thuyết:
 - ❑ $H_0: \delta_1^2 = \delta_2^2$ (phương sai của biến X bằng phương sai của biến Y)
 - ❑ Đối thuyết $H_1: \delta_1^2 \neq \delta_2^2$ với mức ý nghĩa α trong trường hợp kiểm định hai phía. Nếu kiểm định một phía đối thuyết H_1 là: $\delta_1^2 > \delta_2^2$

Vào Data/Data Analysis

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	x	3.4	3.4	3.5	3.5	3.5	3.7	3.7	3.7	3.7	3.9		
3	y	3.2	3.2	3.4	3.4	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.6
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													

Ch4 - Phân tích phương sai, so sánh và kiểm định **31**

Hộp thoại F-Test Two Sample for Variance

Miền vào của mẫu 1, kể cả tên dòng đầu

Miền vào của mẫu 2, kể cả tên dòng đầu

Nhấn đầu dòng

Mức ý nghĩa 0.05 hoặc 0.01

Miền để kết quả

Ch4 - Phân tích phương sai, so sánh và kiểm định **32**



Kết quả

F-Test Two-Sample for Variances		
	x	y
Mean	3.6	3.5
Variance	0.026667	0.025455
Observations	10	12
df	9	11
F	1.047619	
P(F<=f) one-tail	0.462995	
F Critical one-tail	2.896223	

Ch4 - Phân tích phương sai, so sánh và kiểm định

33



Phân tích kết quả

- ❑ Bảng kết quả gồm:
 - ❑ Trung bình của hai mẫu
 - ❑ Phương sai của hai mẫu
 - ❑ Số quan sát n1 và n2: 10 và 12
 - ❑ Bậc tự do: 9 và 11
 - ❑ Giá trị F (thực nghiệm): 1.047619
 - ❑ Giá trị P one-tail (P một phía): 0.46299
 - ❑ Giá trị F Critical one-tail (F lý thuyết một phía): 2.8962

Ch4 - Phân tích phương sai, so sánh và kiểm định

34



Phân tích kết quả

- ❑ Với $F < 1$ nếu $F > F_{\text{Critical one-tail}}$ thì chấp nhận H_0 ($\sigma_1^2 = \sigma_2^2$) ngược lại bác bỏ H_0 , chấp nhận H_1 $\sigma_1^2 > \sigma_2^2$
- ❑ Với $F \geq 1$ nếu $F < F_{\text{Critical one-tail}}$ thì chấp nhận H_0 ($\sigma_1^2 = \sigma_2^2$) ngược lại bác bỏ H_0 , chấp nhận H_1 $\sigma_1^2 > \sigma_2^2$
- ❑ $F = 1.047619 > 1$ và $F < F_{\text{Critical one-tail}} = 2.89622$ nên chấp nhận H_0 ($\sigma_1^2 = \sigma_2^2$)



3. So sánh trung bình

- ❑ Dùng so sánh trung bình hai mẫu thông qua việc kiểm định giả thuyết:
 - ❑ $H_0: m_1 = m_2$ (kỳ vọng biến X bằng kỳ vọng biến Y)
 - ❑ Đối thuyết $H_1: m_1 \neq m_2$ ở mức ý nghĩa α trong trường hợp kiểm định hai phía. Nếu kiểm định một phía thì đối thuyết H_1 là $m_1 > m_2$
- ❑ Có bốn dạng
 - ❑ So sánh hai mẫu độc lập khi biết phương sai σ_1^2, σ_2^2 .
 - ❑ So sánh hai mẫu kiểu cặp đôi.
 - ❑ So sánh hai mẫu độc lập với giả thuyết hai phương sai bằng nhau.
 - ❑ So sánh hai mẫu độc lập với giả thuyết hai phương sai khác nhau.



So sánh cặp đôi và độc lập

Hai mẫu độc lập	Hai mẫu cặp đôi
□ Bố trí dữ liệu ngẫu nhiên (số lần quan sát có thể khác nhau) □ Hai đối tượng khác nhau, với 1 điều kiện môi trường	□ Số quan sát phải bằng nhau □ Một đối tượng, với 2 điều kiện môi trường khác nhau.

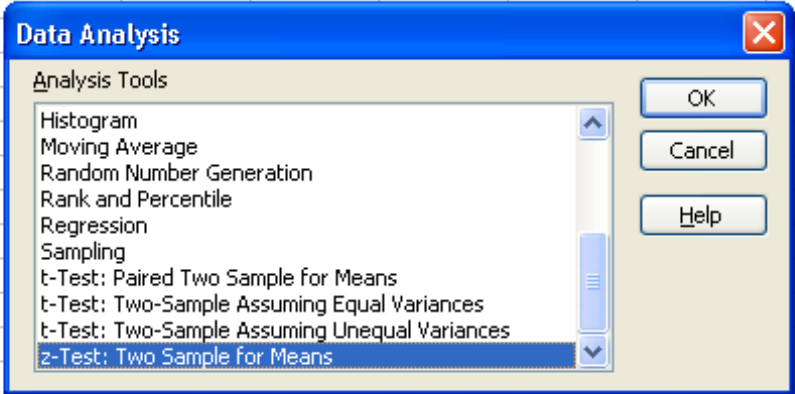


3.1 So sánh hai mẫu độc lập khi biết phương sai σ_1^2, σ_2^2

- **Rút mẫu độc lập từ hai tổng thể** phân phối chuẩn, trong một số tình huống nào đó chúng ta có thể ước lượng được phương sai ví dụ như khi điều tra lại một tổng thể sau một thời gian chưa lâu, nên phương sai chưa thay đổi, do đó lấy phương sai của lần làm trước, từ đó ước lượng về phương sai.

 **Vào Data/Data Analysis**

	C	D	E	F	G	H	I
1	3.4	3.5	3.5	3.5	3.7	3.7	
2	3.2	3.4	3.4	3.6	3.6	3.6	



Ch4 - Phân tích phương sai, so sánh và kiểm định **39**

 **Hiện của sổ**

Miền biến 1 →

Miền biến 2 →

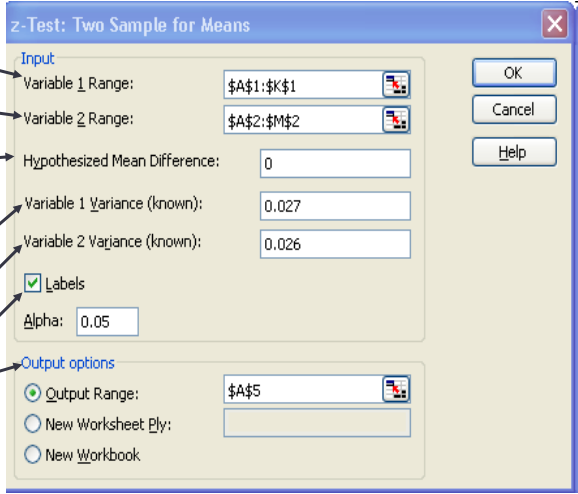
Giả thiết về sự khác nhau của hai trung bình →

Phương sai của biến 1 →

Phương sai của biến 2 →

Nếu có nhãn thì chọn →

Nơi để kết quả →



Ch4 - Phân tích phương sai, so sánh và kiểm định **40**

Kết quả

z-Test: Two Sample for Means			
		x	y
Trung bình	Mean	3.6	3.5
Phương sai	Known Variance	0.027	0.026
Số quan sát	Observations	10	12
Giả thiết sự khác nhau của hai trung bình	Hypothesized Mean Difference	0	
Z thực nghiệm	z	1.433455	
P một phía và hai phía	P(Z<=z) one-tail	0.075864	
	z Critical one-tail	1.644854	
Z lý thuyết (tới hạn) một phía và hai phía	P(Z<=z) two-tail	0.151728	
	z Critical two-tail	1.959964	

Ch4 - Phân tích phương sai, so sánh và kiểm định

41



Phân tích kết quả

- ❑ Z thực nghiệm có thể tính theo công thức:
 - ❑ $Z = (m_1 - m_2) / \sqrt{\delta_1^2 / n_1 + \delta_2^2 / n_2}$
- ❑ P một phía:
 - ❑ $= 1 - \text{Normsdist}(z)$
- ❑ P hai phía
 - ❑ $= (1 - \text{Normsdist}(z)) * 2$
- ❑ Z lý thuyết một phía
 - ❑ $= \text{Normsinv}(0.95)$
- ❑ Z lý thuyết hai phía
 - ❑ $= \text{Normsinv}(0.975)$

Ch4 - Phân tích phương sai, so sánh và kiểm định 42



Phân tích kết quả

□ Các bước phân tích kết quả

□ Trước tiên so sánh Z_{tn} với Z_{lt} hai phía

- Nếu $|Z_{tn}| < Z_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là không khác nhau (chấp nhận giả thuyết H_0)
- Nếu $|Z_{tn}| > Z_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là khác nhau (chấp nhận giả thuyết H_1). Muốn xem trung bình biến nào lớn hơn có thì dùng các cách sau:
 - C1: $Z_{tn} > 0$ thì kết luận trung bình biến 1 lớn hơn và ngược lại
 - C2: $|Z_{tn}| > Z_{lt}$ một phía kết luận $m_1 > m_2$ và ngược lại



Phân tích kết quả

- Ví dụ Z thực nghiệm = 1.43355 < Z lt hai phía = 1.95996 nên kỳ vọng của hai biến không khác nhau (chấp nhận giả thuyết H_0).



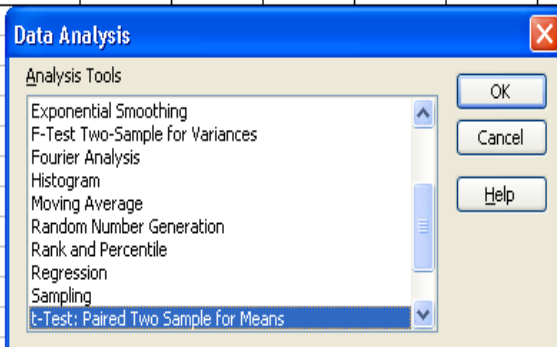
3.2 So sánh hai mẫu cặp đôi

- ❑ Xét một số ví dụ
 - ❑ So sánh trọng lượng giữa các con đực và con cái các đàn lợn giống, lấy ngẫu nhiên lợn đực và lợn cái từng cặp trong các đàn vậy là độc lập hay cặp đôi?
 - ❑ Phân tích mẫu đất?
 - ❑ So sánh các giống lúa?
- ❑ Yêu cầu: Hai mẫu phải cùng số quan sát n , các cặp số liệu xếp thành cặp đứng ở hai cột cạnh nhau.



Vào *Data/Data Analysis*

	A	B	C	D	E	F	G	H	I
1	x	39.6	32.4	33.1	27	36	32	25.9	32.4
2	y	39.2	33.1	32.4	25.2	33.1	29.5	24.1	29.2
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									



Hiện ra của sổ

Miền của biến 1, kể cả hàng đầu của mẫu quan sát.

Miền của biến 2

Giả thiết về hiệu hai trung bình của hai tổng thể. $H_0: m_1 = m_2$ thì ghi 0. Nếu $H_0: m_1 = m_2 + d$ thì ghi d

Nếu có nhãn thì chọn

Chọn miền đặt kết quả

Ch4 - Phân tích phương sai, so sánh và kiểm định **47**

Kết quả

t-Test: Paired Two Sample for Means			
	x	y	Giải thích
Mean	32.3	30.725	Trung bình mẫu
Variance	19.54	23.45071	Phương sai mẫu
Observations	8	8	Số quan sát
Pearson Correlation	0.9618692		Hệ số tương quan R
Hypothesized Mean Difference	0		Giả thuyết về hiệu số trung bình
df	7		Bậc tự do
t Stat	3.3105533		Giá trị t thực nghiệm
P(T<=t) one-tail	0.0064659		Giá trị P một phía
t Critical one-tail	1.8945786		Giá trị t lý thuyết khi kiểm định một phía
P(T<=t) two-tail	0.0129317		Giá trị P hai phía
t Critical two-tail	2.3646243		Giá trị t lý thuyết khi kiểm định hai phía

Ch4 - Phân tích phương sai, so sánh và kiểm định **48**



Phân tích kết quả

- ❑ Ví dụ: $T_{tn} = 3.3 > T_{lt}$ hai phía $\rightarrow$ kỳ vọng của hai biến là khác nhau (chấp nhận giả thuyết H_1)
 - ❑ Cách 1: $T_{tn} > 0$ vậy trung bình $m_x > m_y$
 - ❑ Cách 2: $T_{tn} > T_{lt}$ một phía nên $m_x > m_y$



Phân tích kết quả

- ❑ Các bước phân tích kết quả
 - ❑ Trước tiên so sánh T_{tn} với T_{lt} hai phía
 - ❑ Nếu $|T_{tn}| < T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là không khác nhau (chấp nhận giả thuyết H_0)
 - ❑ Nếu $|T_{tn}| > T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là khác nhau (chấp nhận giả thuyết H_1). Muốn xem trung bình biến nào lớn hơn có thì dùng các cách sau:
 - ❑ C1: $T_{tn} > 0$ thì kết luận trung bình biến 1 lớn hơn và ngược lại
 - ❑ C2: $|T_{tn}| > T_{lt}$ một phía kết luận $m_1 > m_2$ và ngược lại



3.3 So sánh hai mẫu độc lập với giả thuyết hai phương sai bằng nhau

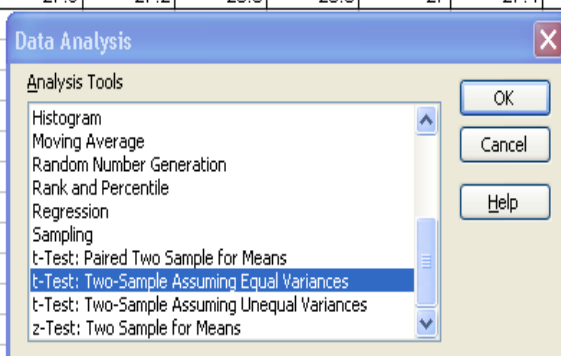
□ Hai mẫu độc lập:

- Nếu dung lượng mẫu lớn (≥ 30) thì ta có thể tiến hành Z-test nhưng thay hai phương sai tổng thể δ_1^2, δ_2^2 bằng phương sai mẫu s_1^2, s_2^2
- Nếu dung lượng mẫu bé (< 30) thì ta gặp bài toán khó (Berens-Fisher)
 - Nếu coi hai phương sai bằng nhau thì dùng *t-Test: Two-Sample Assuming Equal ...*
 - Nếu coi hai phương sai không bằng nhau thì dùng *t-Test: Two-Sample Assuming Unequal ...*



3.3 So sánh hai mẫu độc lập với giả thuyết hai phương sai bằng nhau

	A	B	C	D	E	F	G	H	I
1	x	27.5	27	27.3	27.6	27.8			
2	y	27.9	27.2	26.5	26.3	27	27.4	27.3	26.8
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									





Hộp thoại xuất hiện

Miền biến 1

Miền biến 2

Giả thuyết sự khác nhau hai trung bình

Nếu có nhãn thì chọn

Nơi đặt kết quả

Ch4 - Phân tích phương sai, so sánh và kiểm định

53



Kết quả

Trung bình

Phương sai

Số quan sát

Phương sai chung

Giả thiết sự khác nhau của hai trung bình

Bậc tự do = $n_1 + n_2 - 2$

t thực nghiệm

P một phía và hai phía

t lý thuyết (tới hạn) một phía và hai phía

t-Test: Two-Sample Assuming Equal Variances			
		x	y
Mean		27.44	27.05
Variance		0.093	0.265714
Observations		5	8
Pooled Variance		0.202909	
Hypothesized Mean Difference		0	
df		11	
t Stat		1.518701	
P(T<=t) one-tail		0.078522	
t Critical one-tail		1.795885	
P(T<=t) two-tail		0.157043	
t Critical two-tail		2.200985	

Ch4 - Phân tích phương sai, so sánh và kiểm định

54



Phân tích kết quả

❑ Phương sai chung tính theo CT:

$$s^2 = ((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2) / (n_1 + n_2 - 2)$$

❑ t thực nghiệm

$$t = (m_1 - m_2) / \sqrt{s^2 (1/n_1 + 1/n_2)}$$

❑ Có thể tìm được:

$$P \text{ một phía} = T_{\text{dist}}(z, f, 1)$$

$$P \text{ hai phía} = T_{\text{dist}}(z, f, 2)$$

$$T \text{ lý thuyết một phía} = T_{\text{INV}}(0.1, f)$$

$$T \text{ lý thuyết hai phía} = T_{\text{INV}}(0.05, f)$$



Phân tích kết quả

❑ Các bước phân tích kết quả

❑ Trước tiên so sánh T_{tn} với T_{lt} hai phía

❑ Nếu $|T_{tn}| < T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là không khác nhau (chấp nhận giả thuyết H_0)

❑ Nếu $|T_{tn}| > T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là khác nhau (chấp nhận giả thuyết H_1). Muốn xem trung bình biến nào lớn hơn có thì dùng các cách sau:

❑ C1: $T_{tn} > 0$ thì kết luận trung bình biến 1 lớn hơn và ngược lại

❑ C2: $|T_{tn}| > T_{lt}$ một phía kết luận $m_1 > m_2$ và ngược lại



Phân tích kết quả

- ❑ Ví dụ: $t_{tn} = 1.5187 < t_{lt \text{ hai phía}} = 2.20099$ (ứng với $P \text{ một phía}$ và $P \text{ hai phía}$) nên chấp nhận giả thuyết H_0 ở mức ý nghĩa $\alpha = 0.05$



Chú ý

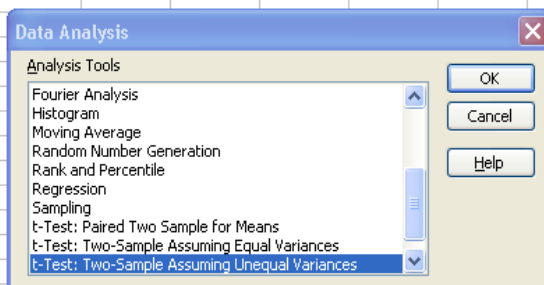
- ❑ Trước khi dùng mô hình này cần kiểm định giả thuyết phụ $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 > \sigma_2^2$ có thể tính toán bằng phân phối Fisher:
 - ❑ B1: Tính F thực nghiệm bằng tỷ số giữa hai phương sai mẫu (lấy lớn chia cho bé)
 - ❑ B2: So F thực nghiệm với F lý thuyết ở mức ý nghĩa α : $= \text{FINV}(\alpha, f_1, f_2)$. Nếu $F_{tn} < F_{lt}$ thì chấp nhận H_0 ngược lại chấp nhận H_1
- ❑ Ví dụ trên: $F_{tn} = 0.26571 / 0.093 = 2.8571$ và $F_{lt} = \text{FINV}(0.05, 7, 4) = 6.09421$ nên chấp nhận H_0



3.4 So sánh hai mẫu độc lập với giả thuyết hai phương sai không bằng nhau

- Trong trường hợp mẫu bé và phương sai khác nhau

	A	B	C	D	E	F	G	H	I
1	x	27.5	27	27.3	27.6	27.8			
2	y	27.9	27.2	26.5	26.3	27	27.4	27.3	26.8
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									



Ch4 - Phân tích phương sai, so sánh và kiểm định

59



Hộp thoại xuất hiện

t-Test: Two-Sample Assuming Unequal Variances

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☒ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

Ch4 - Phân tích phương sai, so sánh và kiểm định

60



Kết quả

5	t-Test: Two-Sample Assuming Unequal Variances		
6			
7		x	y
8	Mean	27.44	27.05
9	Variance	0.093	0.265714
10	Observations	5	8
11	Hypothesized Mean Difference	0	
12	df	11	
13	t Stat	1.713325	
14	P(T<=t) one-tail	0.057329	
15	t Critical one-tail	1.795885	
16	P(T<=t) two-tail	0.114657	
17	t Critical two-tail	2.200985	

Ch4 - Phân tích phương sai, so sánh và kiểm định

61



Phân tích kết quả

□ Các bước phân tích kết quả

□ Trước tiên so sánh T_{tn} với T_{lt} hai phía

□ Nếu $|T_{tn}| < T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là không khác nhau (chấp nhận giả thuyết H_0)

□ Nếu $|T_{tn}| > T_{lt}$ hai phía (ở mức ý nghĩa α) thì kết luận kỳ vọng của hai biến là khác nhau (chấp nhận giả thuyết H_1). Muốn xem trung bình biến nào lớn hơn có thì dùng các cách sau:

□ C1: $T_{tn} > 0$ thì kết luận trung bình biến 1 lớn hơn và ngược lại

□ C2: $|T_{tn}| > T_{lt}$ một phía kết luận $m_1 > m_2$ và ngược lại

Ch4 - Phân tích phương sai, so sánh và kiểm định

62



Phân tích kết quả

- ❑ t thực nghiệm = 1.71 < t lý thuyết hai phía = 2.2 (P một phía và hai phía đều lớn hơn mức ý nghĩa $\alpha = 0.05$) nên chấp nhận giả thuyết H_0