

PHÂN PHỐI VÀ PHÉP KIỂM KHI BÌNH PHƯƠNG

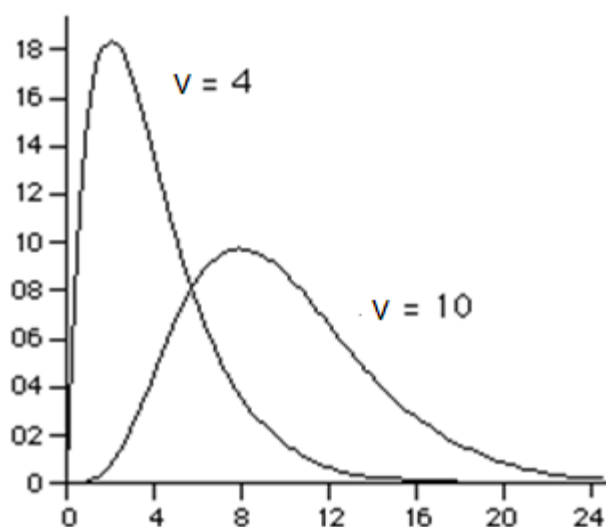
12.1 Phân phối khi bình phương (χ^2):

Nếu X_i là biến ngẫu nhiên có phân phối chuẩn tắc $X_i \sim N(0,1)$ và gọi:

$$Q = \sum_{i=1}^v X_i^2$$

thì Q tuân theo luật phân phối khi bình phương (χ^2) với bậc tự do v và được viết là:

$$Q \sim \chi_v^2$$



Biểu đồ 12.1 Phân phối χ^2 với bậc tự do $v=4$ và $v=10$.

Phân phối (χ^2) có các đặc điểm:

- Trung bình của phân phối chính là bậc tự do v ($\mu = v$).
- Phương sai bằng 2 lần bậc tự do: $\sigma^2 = 2v$
- y có trị cực đại (mode) khi $x^2 = v-2$ (với $v \geq 2$).
- Khi bậc tự do v tăng, đường cong khi bình phương tiến dần thành đường cong chuẩn.

Như vậy phân phối χ^2 chỉ có 1 thông số là bậc tự do v

Ví dụ: bậc tự do $v=4$, trị trung bình là 4, phương sai là 8, trị cực đại là 2

bậc tự do $v=10$, trị trung bình là 10, phương sai là 20, trị cực đại là 8

12.2 Sử dụng phép kiểm χ^2

Thường được dùng để kiểm định các biến định tính.

12.2.1 Kiểm định sự xác hợp (Goodness of fit) của 1 tần số quan sát trên mẫu, so sánh với tần số lý thuyết.

Ví dụ: Tần số lý thuyết (hay tần số mong đợi) sinh trai hoặc gái là 50%. Quan sát một mẫu: Số trẻ em sinh tại khoa Sản trong 1 tuần lễ gồm 110 trẻ trai và 90 trẻ gái (tần số quan sát). Như vậy có thể kết luận hiện nay có khuynh hướng sinh trai nhiều hơn gái?

Giả thuyết H_0 : tần số quan sát=tần số lý thuyết

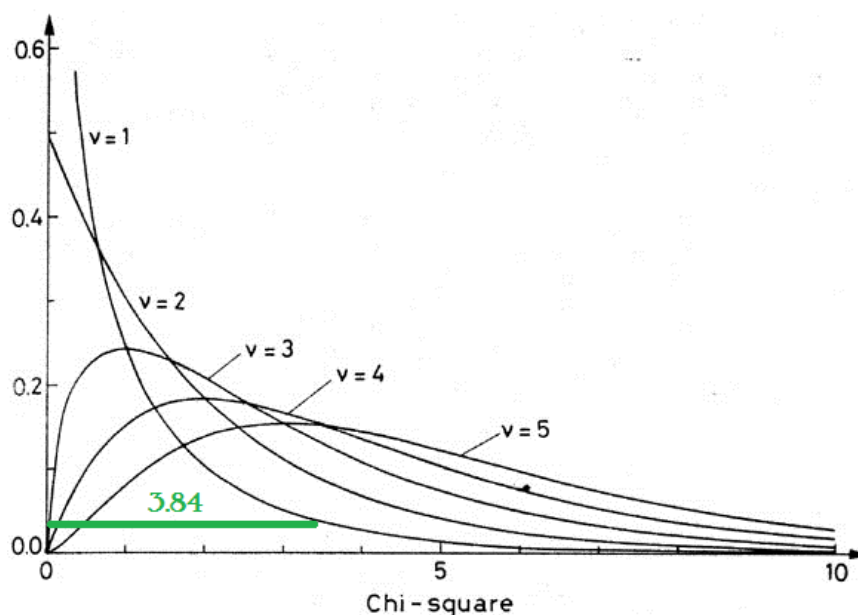
Giả thuyết H_a : Có khác biệt

$$\chi^2 = \sum \frac{(\text{TS quan sát} - \text{TS mong đợi})^2}{\text{TS mong đợi}}$$

$$\text{Tính } \chi^2: \chi^2 = \frac{(110 - 100)^2 + (90 - 100)^2}{100} = 2$$

Xem bảng phân phối χ^2 (phân phụ lục) với bậc tự do $DF=1$, giá trị tới hạn $\alpha=0,05$ là 3,84

$2 < 3,84 \rightarrow$ chấp nhận H_0 : Sinh trai nhiều hơn gái là do tình cờ vì sự khác biệt chưa có ý nghĩa thống kê. (xem biểu đồ 12.2)



Biểu đồ 12.2 Với bậc tự do $v=1$, $\chi^2 = 3,84$

12.2.2 Kiểm định 2 tỉ lệ trên 2 mẫu độc lập (Bảng 2x2)

Áp dụng khi giá trị mong đợi trung bình trong các ô ≥ 5

Ví dụ: Một nghiên cứu bệnh-chứng tìm sự liên hệ giữa nhiễm Human papilloma virus (HPV) và ung thư cổ tử cung (CTC). Kết quả trình bày trong bảng 2x2 như sau:

	Ung thư CTC+	Ung thư CTC(-)	
HPV (+)	20 (10)*	10 (20)*	30
HPV (-)	10 (20)*	50 (40)*	60
	30	60	N=90**

*Các số trong ngoặc là số mong đợi, chỉ áp dụng phép kiểm khi trị số mong đợi ≥ 5 .

** N là tổng số ca

Cách tính khi bình phương:

$$\chi^2 = \sum \frac{(\text{trị số quan sát} - \text{trị số mong đợi})^2}{\text{số mong đợi}} \quad [1]$$

$$\text{Trị số mong đợi} = \frac{\text{tổng hàng} \times \text{tổng cột}}{N} \quad [2]$$

Bậc tự do DF = (hàng-1) x (cột-1)

Bảng trên có 2 hàng và 2 cột vậy:

$$DF = (2-1) \times (2-1) = 1$$

Thế vào công thức [1] ta có

$$\chi^2 = \frac{(20-10)^2}{10} + \frac{(10-20)^2}{20} + \frac{(10-20)^2}{20} + \frac{(50-40)^2}{40} = 22,5$$

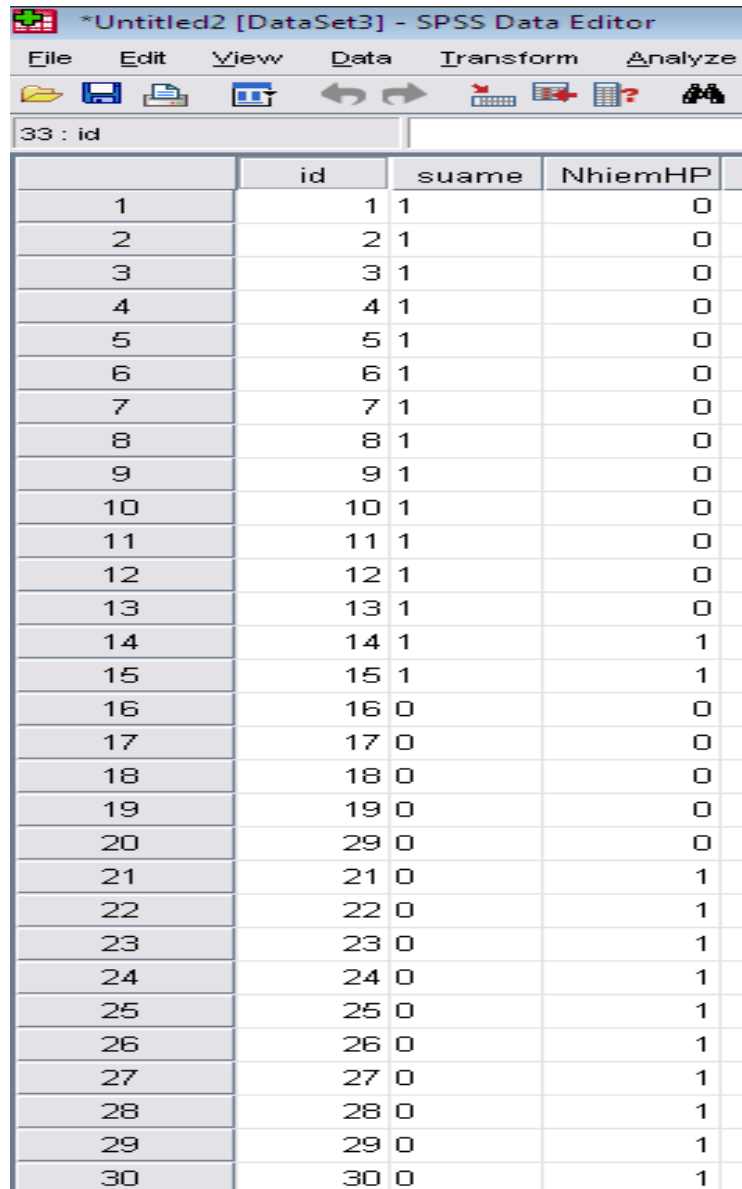
Với DF=1; giá trị tới hạn với $\alpha=0,05$ (2 đuôi) là 3,84

$\chi^2 = 22,5 > 3,84 \rightarrow$ bác bỏ H_0 : nhiễm HPV là yếu tố nguy cơ gây ung thư cổ tử cung

12.3 Áp dụng phép kiểm χ^2 trong SPSS

12.3.1 Trường hợp có bảng dữ liệu:

Ví dụ: Một nghiên cứu tìm sự liên quan giữa bú sữa mẹ và nhiễm *Helicobacter pylori* ở 30 trẻ em. Ta có bảng dữ liệu sau đây:



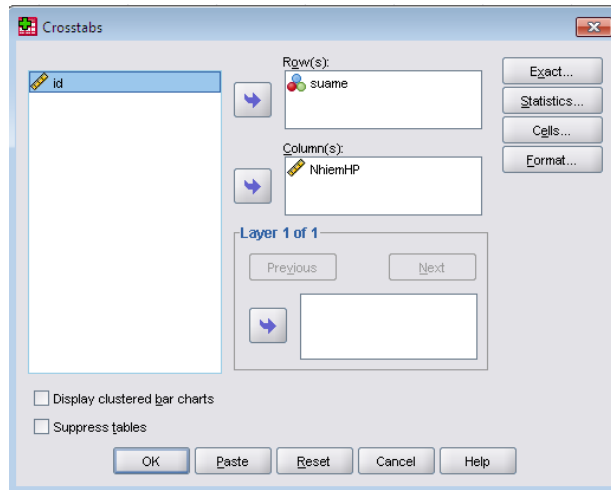
The screenshot shows the SPSS Data Editor window titled '*Untitled2 [DataSet3] - SPSS Data Editor'. The menu bar includes File, Edit, View, Data, Transform, and Analyze. Below the menu bar is a toolbar with various icons. The main window displays a data grid with 30 rows and 4 columns. The columns are labeled 'id', 'suame', and 'NhiemHP'. The first column (id) contains values from 1 to 30. The second column (suame) contains values 1 or 0. The third column (NhiemHP) contains values 0 or 1.

	id	suame	NhiemHP
1	1	1	0
2	2	1	0
3	3	1	0
4	4	1	0
5	5	1	0
6	6	1	0
7	7	1	0
8	8	1	0
9	9	1	0
10	10	1	0
11	11	1	0
12	12	1	0
13	13	1	0
14	14	1	1
15	15	1	1
16	16	0	0
17	17	0	0
18	18	0	0
19	19	0	0
20	29	0	0
21	21	0	1
22	22	0	1
23	23	0	1
24	24	0	1
25	25	0	1
26	26	0	1
27	27	0	1
28	28	0	1
29	29	0	1
30	30	0	1

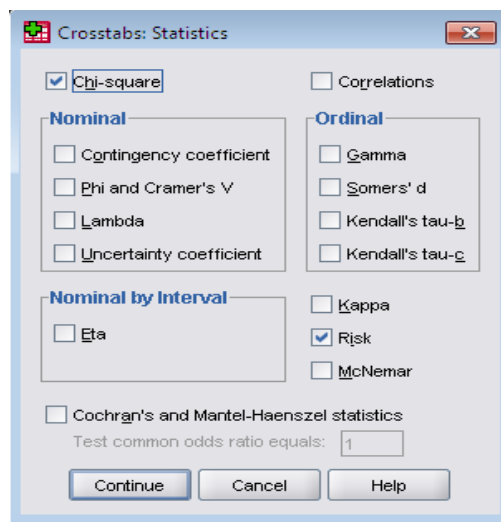
Ghi chú: cột 1 (id): đối tượng; cột 2 (suame) với 1: bú sữa mẹ và 0: không bú mẹ ;
cột 3 (nhiemHP): 1 có nhiễm *H. pylori* và 0: không nhiễm *H. pylori*

Vào menu: **Analyze > Descriptive Statistics > Crosstabs.**

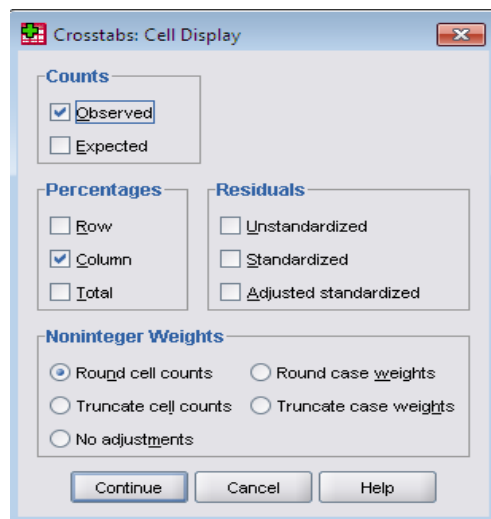
Mở hộp thoại Crosstabs, nhấp chuyển suame từ ô bên trái vào ô Row(s) và NhiemHP vào ô column(s).



Nhấn hộp thoại Statistics...Nhấp dấu nháy ✓ vào ô Chi-square và ô Risk và nhấn Continue



Nhấn hộp thoại Cells ...Nhấp dấu nháy ✓ vào ô Observed và ô Column (để xem tỉ lệ %), nhấp Continue



Cuối cùng nhấn OK cho kết quả như sau:

suame * NhiemHP Crosstabulation

			NhiemHP		Total
			0	1	
suame	0	Count	5	10	15
		% within NhiemHP	27.8%	83.3%	50.0%
	1	Count	13	2	15
		% within NhiemHP	72.2%	16.7%	50.0%
Total		Count	18	12	30
		% within NhiemHP	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.889 ^a	1	.003		
Continuity Correction ^b	6.806	1	.003		
Likelihood Ratio	9.505	1	.002		
Fisher's Exact Test				.008	.004
Linear-by-Linear Association	8.593	1	.003		
N of Valid Cases ^c	30				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.00.

b. Computed only for a 2x2 table

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for suame (0 / 1)	.077	.012	.482
For cohort NhiemHP = 0	.385	.183	.808
For cohort NhiemHP = 1	5.000	1.311	19.074
N of Valid Cases	30		

Kết quả: Không bú sữa mẹ, tỉ lệ nhiễm HP là 10/15 (83,3%)

Có bú sữa mẹ, tỉ lệ nhiễm HP là 2/15 (16,7%)

Pearson Chi-Square=8,889, bậc tự do DF=1 và mức ý nghĩa p=0,003

Tỉ số Odds=0,077 (Khoảng tin cậy 95%: 0,012-0,480)

Kết luận: Bú sữa mẹ làm giảm 92,3% nguy cơ (odds) nhiễm *H. pylori*

12.3.2 Trường hợp không có bảng dữ liệu:

Ví dụ chỉ có bảng dữ liệu 2x2 như sau:

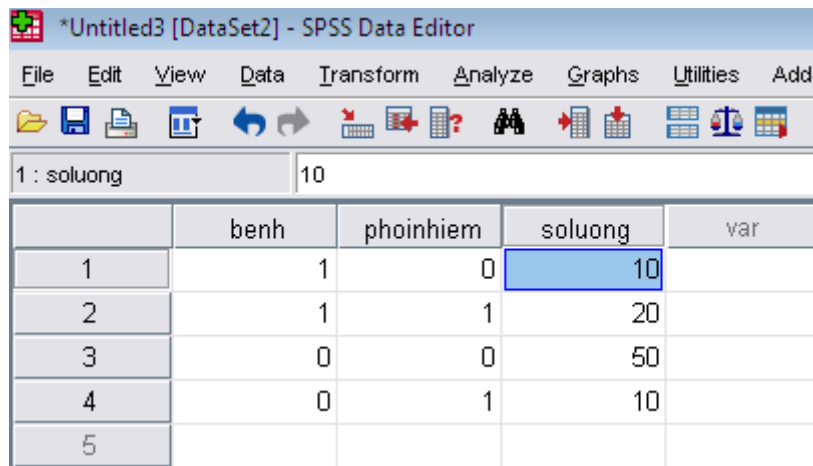
	Ung thư CTC+	Ung thư CTC(-)
HPV (+)	20	10
HPV (-)	10	50

Trong trường hợp này, nhập số liệu vào SPSS như sau:

Có 3 cột: *benh* (0,1); *phoinhiem* (0,1), *soluong* (10,20,50,10)

benh= 1 (ung thư cổ TC +); *benh*=0 (ung thư cổ TC -)

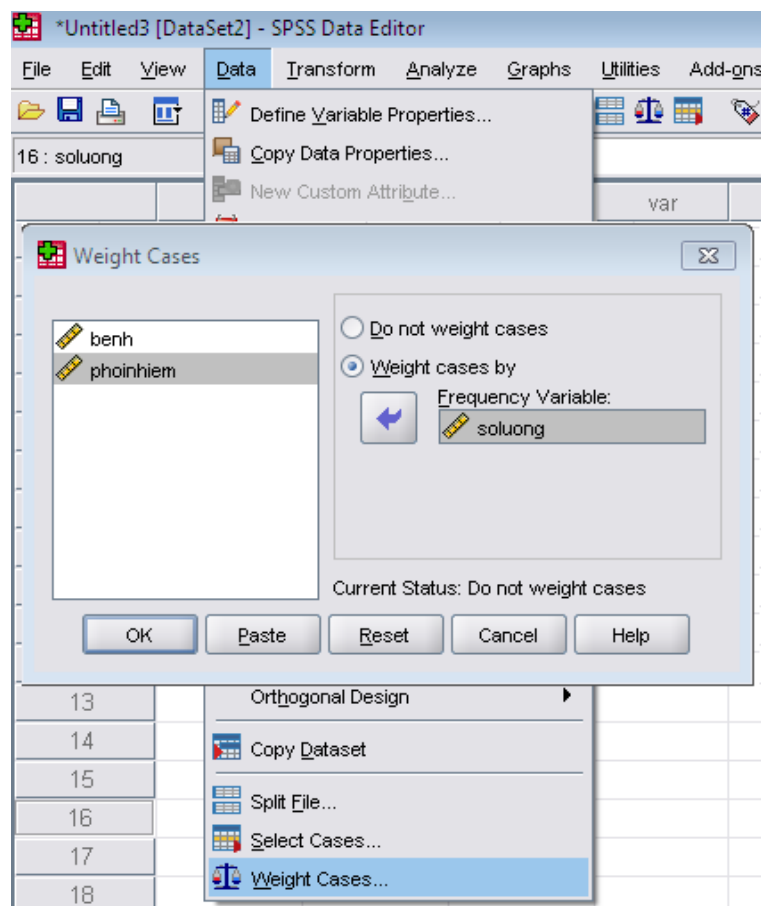
phoinhiem=1 (HPV +); *phoinhiem*=0 (HPV-)



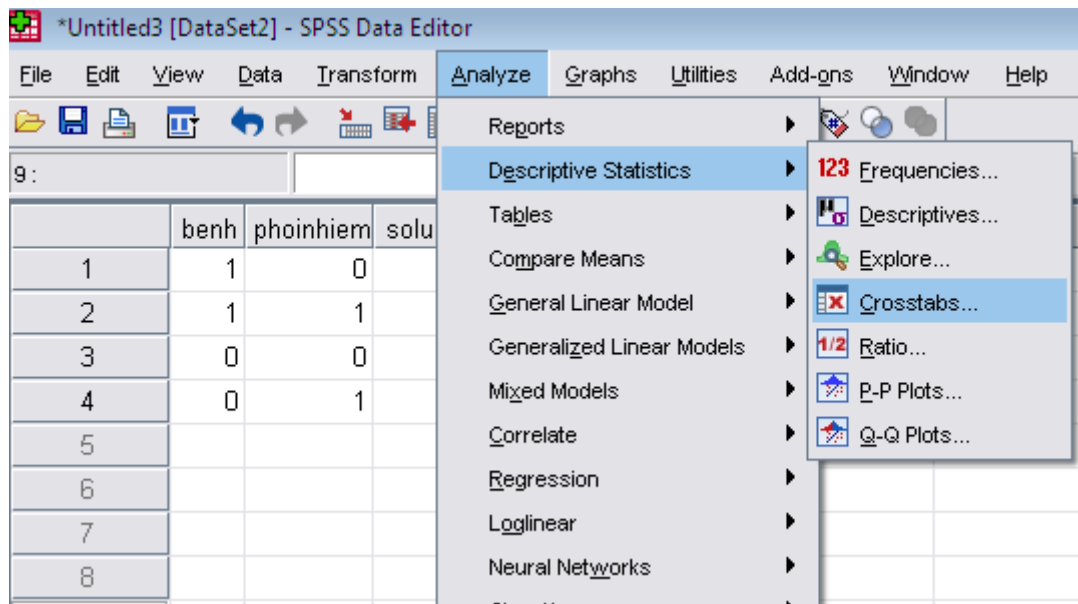
	<i>benh</i>	<i>phoinhiem</i>	<i>soluong</i>	<i>var</i>
1	1	0	10	
2	1	1	20	
3	0	0	50	
4	0	1	10	
5				

Vào menu **Data**, chọn **Weight Cases** (hàng cuối cùng) mở màn hình Weight Cases như sau:

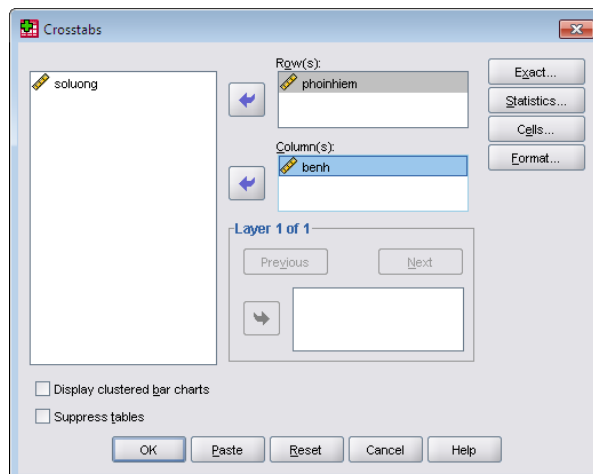
Chọn *soluong*, nhấp mũi tên, chuyển vào ô Frequency Variable, nhấp OK



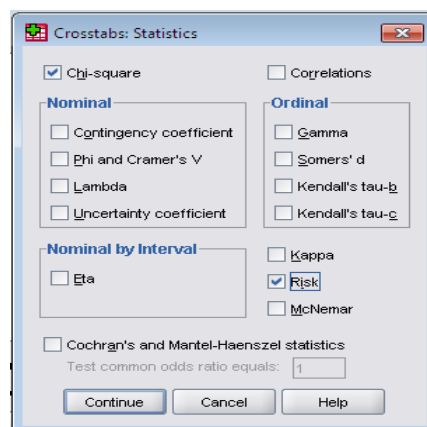
Chuyển qua menu: **Analyze**, chọn **Descriptive Statistics**, chọn **Crosstabs**



Vào hộp thoại Cross tabs, nhấp chuyển phoinhiem từ ô bên trái qua ô Row(s) và bệnh qua ô Column(s):



Nhấn Statistics, đánh dấu nháy ✓ vào ô Chi-square và Risk



Nhấp Continue, nhấp OK, kết quả cuối cùng như sau:

phoinhiem *benh Crosstabulation

Count		benh		Total
		0	1	
phoinhiem	0	50	10	60
	1	10	20	30
Total		60	30	90

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	22.500 ^a	1	.000	.000
Continuity Correction ^b	20.306	1	.000	
Likelihood Ratio	22.314	1	.000	
Fisher's Exact Test				
Linear-by-Linear Association	22.250	1	.000	
N of Valid Cases ^b	90			

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 10.

b. Computed only for a 2x2 table

Risk Estimate

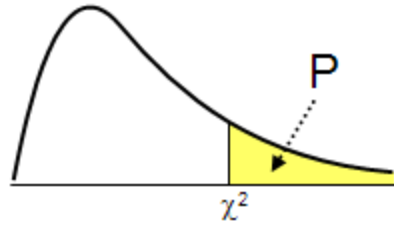
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for phoinhiem (0 / 1)	10.000	3.612	27.688
For cohort benh = 0	2.500	1.488	4.199
For cohort benh = 1	.250	.135	.465
N of Valid Cases	90		

Pearson Chi-Square = 22.500; Ý nghĩa TK (2-đuôi)=0.000

Odds Ratio=10,0 (KTC 95%:3,6-27,6)

Kết luận: Nguy cơ mắc bệnh ung thư cổ tử cung trong nhóm nhiễm HPV cao hơn nhóm không nhiễm HPV 10 lần (KTC 95%: 3.612- 27.688; p=0.000).

Phụ lục. Bảng đối chiếu xem giá trị tới hạn với 6 bậc tự do (DF) của phân phối χ^2 .



DF	p=0.05	p=0.01
1	3.84	6.64
2	5.99	9.21
3	7.82	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81