không có cơ sở lý luận rõ rệt; nếu để tuỳ ý chia thì có người đem d_x chia cho d_y, nhưng cũng có người đem d_y chia cho d_x, kết quả sẽ không thống nhất. Do đó dùng phương pháp chia để tập hợp cả d_x và d_y thành một chữ số cũng không được.

Chỉ có dùng phương pháp nhân để tập hợp các d_x và d_y thành một chữ số mới giải quyết được các vấn đề trên và đó chính là hệ số tương quan:

$$r = \frac{\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)}{n-1} \quad (11\text{-}31)$$

$$= \frac{1}{s_x . s_y}\left[\frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}\right] = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 . \sum(y-\bar{y})^2}}$$ [(1)]

Nếu vẽ biểu đồ điểm tương quan và các góc tương quan I, II, III và IV, rồi phân tích dấu và trị số tuyệt đối của tích $(x - \bar{x})(y - \bar{y})$ thấy:

Khi biến số x và biến số y hoàn toàn tương quan thuận, tức trị số x lớn sẽ cùng trị số y lớn xếp thành một cặp, trị số x nhỏ sẽ cùng trị số y nhỏ xếp thành một cặp, cũng tức là $(x - \bar{x})$ có dấu dương cùng $(y - \bar{y})$ có dấu dương sẽ xếp thành một cặp, và $(x - \bar{x})$ có dấu âm cùng $(y - \bar{y})$ có dấu âm xếp thành một cặp thì tích $(x - \bar{x})(y - \bar{y})$ luôn luôn mang dấu dương. Như vậy hệ số tương quan r cũng mang dấu dương và có trị số tuyệt đối lớn nhất. Trường hợp đó các điểm tương quan đều nằm trên 1 đường thẳng ở góc I và góc III.

Khi biến số x và biến số y hoàn toàn tương quan nghịch, tức trị số x lớn sẽ cùng trị số y nhỏ xếp thành một cặp, trị số x nhỏ sẽ cùng trị số y lớn xếp thành một cặp, cũng tức là $(x - \bar{x})$ có dấu dương cùng $(y - \bar{y})$ có dấu âm xếp thành một cặp và $(x - \bar{x})$ có dấu âm cùng $(y - \bar{y})$ có dấu dương xếp thành một cặp, thì tích $(x - \bar{x})(y - \bar{y})$ luôn luôn mang dấu âm. Như vậy hệ số tương quan r cũng mang dấu âm và có trị số tuyệt đối lớn nhất. Trường hợp đó các điểm tương quan đều nằm trên 1 đường thẳng ở góc II và IV.

[(1)] Trong đó tỉ số:

$$= \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$

gọi là hiệp phương sai, ký hiệu là Cov (x, y), tức:

$$Cov(x,y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{n-1}$$

Khi Cov(x, y) = 0 biểu thị x và y không có quan hệ tương quan.

Giá trị của hiệp phương sai có thể dương, có thể âm. Hiệp phương sai mang dấu dương biểu thị x và y tương quan thuận, hiệp phương sai mang dấu âm biểu thị x và y tương quan nghịch.

Hiệp phương sai mới chỉ cho ta biết x và y có tương quan với nhau hay không và tương quan thuận hay nghịch, hiệp phương sai không cho ta biết mức độ chặt chẽ của tương quan.

Khi biến số x và biến số y không hoàn toàn tương quan thuận, tức là lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu dương, lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu âm, nhưng đại đa số tích $(x - \bar{x})(y - \bar{y})$ mang dấu dương, do đó hệ số tương quan vẫn mang dấu dương và có trị số tuyệt đối nhỏ hơn. Trường hợp này đại đa số các điểm tương quan nằm ở góc I và góc III, một số ít điểm tương quan nằm ở góc II và góc IV; các điểm tương quan này họp thành một hình bầu dục nghiêng về phải.

Khi biến số x và biến số y không hoàn toàn tương quan nghịch, tức là lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu dương, lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu âm, nhưng đại đa số tích $(x - \bar{x})(y - \bar{y})$ mang dấu âm, do đó hệ số tương quan vẫn mang dấu âm và có trị số tuyệt đối nhỏ hơn. Trường hợp này đại đa số các điểm tương quan nằm ở góc II và góc IV, một số ít điểm tương quan nằm ở góc I và góc III; các điểm tương quan này họp thành một hình bầu dục nghiêng về trái.

Khi các biến số x và biến số y hoàn toàn không tương quan, tức là lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu dương, lúc thì tích $(x - \bar{x})(y - \bar{y})$ mang dấu âm và tổng giá trị của hai loại tích này bằng nhau, do đó hệ số tương quan sẽ bằng không. Trường hợp này các điểm tương quan phân bố đều trên 4 góc I, II, III và IV.

Vì vậy dấu dương âm của hệ số tương quan biểu thị hướng tương quan: hệ số tương quan là dương biểu thị tương quan thuận, hệ số tương quan mang dấu âm biểu thị tương quan nghịch. Còn trị số tuyệt đối của hệ số tương quan biểu thị mức độ tương quan: trị số tuyệt đối càng lớn biểu thị mức độ tương quan càng chặt chẽ, trị số tuyệt đối càng nhỏ biểu thị mức độ tương quan càng không chặt chẽ. Hệ số tương quan biến thiên từ -1 đến +1, hệ số tương quan bằng +1 biểu thị hoàn toàn tương quan thuận, hệ số tương quan bằng -1 biểu thị hoàn toàn tương quan nghịch, hệ số tương quan bằng 0 biểu thị hàn toàn không tương quan.

b. Công thức và thuật toán tính hệ số tương quan đường thẳng (r)

α. Đối với tính trạng định lượng:

- *Trường hợp mẫu nhỏ* (số cặp quan sát hoặc đo lường nhỏ hơn 30)

Công thức thực hành để tính hệ số tương quan như sau:

$$r = \frac{\sum x.y - \frac{\sum x.\sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right]}} \qquad (11\text{-}32)$$

Trong đó: r là hệ số tương quan.

x là các giá trị khác nhau của tính trạng thứ nhất.

y là các giá trị khác nhau của tính trạng thứ hai.

n là số cặp quan sát hoặc đo lường.

Để tính được hệ số tương quan cần lập một bảng tính trạng đó có các cột x, y, xy, x^2 và y^2.

Thí dụ: Xác định hệ số tương quan giữa tỉ lệ mỡ sữa của đời con (y) và đời mẹ x) theo tài liệu ở bảng 11-2.

Lập bảng tính hệ số tương quan trong trường hợp mẫu nhỏ (xem bảng 11-6).

Bảng 11-6. Bảng tính hệ số tương quan trong trường hợp mẫu nhỏ

Số thứ tự	x	y	xy	x^2	y^2
1	3,4	3,5	11,90	11,56	12,25
2	3,8	3,8	14,44	14,44	14,44
3	4,0	3,9	15,60	16,00	15,21
4	3,0	3,3	9,90	9,00	10,89
5	3,7	3,7	13,69	13,69	13,69
6	3,0	3,4	10,20	9,00	11,56
7	3,5	3,6	12,60	12,25	12,96
8	3,8	4,0	15,20	14,44	16,00
9	3,1	3,4	10,54	9,61	11,56
10	3,0	3,1	9,30	9,00	9,61
n = 10	$\Sigma x = 34,3$	$\Sigma y = 35,7$	$\Sigma xy = 123,37$	$\Sigma x^2 = 118,99$	$\Sigma y^2 = 128,17$

Thay các giá trị của Σx, Σy, Σxy, Σx^2, Σy^2 và n vào công thức 11-32 được:

$$r = \frac{123,37 - \dfrac{(34,3)(35,7)}{10}}{\sqrt{\left[118,99 - \dfrac{(34,3)^2}{10}\right]\left[128,17 - \dfrac{(35,7)^2}{10}\right]}} = \pm 0,93$$

- *Trường hợp mẫu lớn* (số cặp quan sát hoặc đo lường bằng hoặc lớn hơn 30).

Công thức thực hành để tính hệ số tương quan như sau:

$$r = \frac{\sum\left\{\left[\sum\left(f_{xy}.a_x\right)a_y\right]\right\} - n.b_x.b_y}{n.s_x.s_y} \qquad (11\text{-}33)$$

Trong đó:

$$b_x = \frac{\sum f_x.a_x}{n} \quad \text{và} \quad b_y = \frac{\sum f_y.a_y}{n} \qquad (11\text{-}34)$$

$$s_x = \sqrt{\frac{\sum f_x a_x^2 - \frac{\left(\sum f_x a_x\right)^2}{n}}{n}} \quad \text{và} \quad s_y = \sqrt{\frac{\sum f_y a_y^2 - \frac{\left(\sum f_y a_y\right)^2}{n}}{n}} \qquad (11\text{-}35)$$

r là hệ số tương quan

f_{xy} là tần số của hàng x và cột y trong bảng phân bố tần số hai chiều (bảng 11-4).

f_x là tần số của các tổ theo x.

f_y là tần số của các tổ theo y.

a_x là trung bình độ lệch của các tổ (theo x) tới tổ có số trung bình giả sử.

a_y là trung bình độ lệch của các tổ (theo y) tới tổ có số trung bình giả sử.

n là số cặp quan sát hoặc đo lường.

Muốn tính hệ số tương quan theo công thức trên cũng phải lập bảng phân bố tần số hai chiều, sau đó lập các cột f_x, a_x, $f_x.a_x$, $f_x.a_x^2$, lập các hàng f_y, a_y, $f_y.a_y$, $f_y.a_y^2$, rồi tính tổng của chúng (xem bảng 10-3). Đồng thời cũng dựa vào bảng này để tính $\sum\left\{\left[\sum\left(f_{xy}.a_x\right)\right]a_y\right\}$.

Thí dụ: Xác định hệ số tương quan giữa trọng lượng và vòng ngực của trâu theo tài liệu ở bảng 1-6.

Tính các giá trị của b_x, b_y, s_x và s_y:

$$b_x = \frac{+30}{100} = +0,30$$

$$b_y = \frac{+22}{100} = +0,22$$

$$s_x = \sqrt{\frac{474}{100} - \left(\frac{30}{100}\right)^2} = 2,156$$

$$s_y = \sqrt{\frac{378}{100} - \left(\frac{22}{100}\right)^2} = 1,931$$

Thay các giá trị của $\Sigma f_{xy}.a_x.a_y$, b_x, b_y, s_x, s_y và n vào công thức 11-33 được:

$$r = \frac{345 - 100.(0,3)(0,22)}{100.(2,156).(1,931)} = +0,812$$

β. Đối với tính trạng định tính (r_a)

Công thức thực hành để tính hệ số tương quan giữa hai tính trạng định tính như sau:

$$r_a = \frac{f_1.f_4 - f_2.f_3}{\sqrt{(f_1 + f_2)(f_3 + f_4)(f_1 + f_3)(f_2 + f_4)}} \qquad (11\text{-}36)$$

Trong đó: r_a là hệ số tương quan giữa hai tính trạng định tính.

f_1, f_2, f_3 và f_4 là các tần số trong mỗi ô của mạng lưới tương quan định tính.

Thí dụ: Xác định hệ số tương quan giữa vấn đề chọn lọc gà theo khả năng miễn dịch đối với bệnh bạch lị và tỉ lệ gà con bị mắc bệnh bạch dị theo tài liệu dưới đây[(1)] (xem bảng 11-7).

Bảng 11-7. Số lượng gà bị mắc bệnh lị ở đàn gà không chọn lọc và có chọn lọc theo khả năng miễn dịch đối với bệnh bạch lị

Gà con	Không mắc bệnh bạch lị	Mắc bệnh bạch lị	Cộng
Ở nhóm không chọn lọc	115 (f_1)	105 (f_2)	$f_1 + f_2 = 220$
Ở nhóm chọn lọc	560 (f_3)	58 (f_4)	$f_3 + f_4 = 618$
Cộng	$f_1 + f_3 = 675$	$f_2 + f_4 = 163$	-

Thay thế các giá trị tính được ở bảng 11-7 vào công thức (11-36) được:

$$r_a = \frac{115 \times 58 - 105 \times 560}{\sqrt{220 \times 618 \times 675 \times 163}} = -0,43$$

γ. Hệ số tương quan giữa tính trạng định lượng và tính trạng định tính (r_b):

Công thức thực hành để tính hệ số tương quan giữa tính trạng định lượng và tính trạng định tính như sau:

$$r_b = \frac{\dfrac{\sum f_+ a}{n_+} - \dfrac{\sum fa}{n}}{\sqrt{C\left(\dfrac{1}{n_+} - \dfrac{1}{n}\right)}} \qquad (11\text{-}37)$$

Trong đó: C: là tổng bình phương:

$$C = \sum fa^2 - \frac{\left(\sum fa\right)^2}{n} \qquad (11\text{-}38)$$

f là tần số của dãy biến thiên, tính theo tính trạng định lượng.

f_+ là tần số của dãy biến thiên, tính theo một tính trạng định tính nào đó.

n là dung lượng quan sát hoặc đo lường tính theo tính trạng định lượng.

n_+ là dung lượng quan sát hoặc đo lường tính theo tính trạng định tính có f_+.

a là trung bình độ lệch của các tổ tới tổ có số trung bình giả sử.

[(1) (3)] Cũng có thể xác định ảnh hưởng của việc chọn lọc gà theo khả năng miễn dịch đối với bệnh bạch lị tới tỉ lệ gà con bị mắc bệnh bạch lị theo phương pháp xác định mức độ sai khác nhau giữa hai số trung bình đối với tính trạng định tính.

Thí dụ: xác định hệ số tương quan giữa trọng lượng sơ sinh của bê và các dạng hê-mô-glô-bin theo tài liệu dưới đây (xem bảng 11-8).

Bảng 11-8. Quan hệ giữa trọng lượng sơ sinh của bê và dạng bê-mô-glô-bin

Trọng lượng sơ sinh (kg) / Dạng hê-mô-glô-bin	19 - 20	21 - 23	24 - 26	27 - 29	30 - 32
Hb.AA (f_+)	10	20	30	40	20
Hb.AB (f_-)	10	20	30	10	10

Muốn tính r_b ta lập bảng tính sau (xem bảng 11-9).

Bảng 11-9. Bảng tính hệ số tương quan giữa tính trạng định lượng và tính trạng định tính

Trọng lượng sơ sinh của bê (kg) / Dạng hê-mô-glô-bin	19-20	21-23	24-26	27-29	30-32	
Hb.AA (f_+)	10	20	30	40	20	$\Sigma f_+ = n_+ = 120$
Hb.AB (f_-)	10	20	30	10	10	$\Sigma f_- = n_- = 80$
f	20	40	60	50	30	$n_+ + n_- = 200$
a	-2	-1	0*	1	2	-
$f_+.a$	-20	-20	0	40	40	$\Sigma f_+a = -40 + 80 = +40$
fa	-40	-40	0	50	60	$\Sigma fa = -80 + 110 = +30$
fa^2	80	40	0	50	120	$\Sigma fa^2 = 290$

* Trong trường hợp này lấy A = 25 là số trung bình giả sử.

Tính tổng bình phương C:

$$C = 290 - \frac{30^2}{200} = 285,5$$

Thay các giá trị của Σf_+a, Σfa, C, n_+ và n vào công thức 11-37 được:

$$r_b = \frac{\frac{40}{120} - \frac{30}{200}}{\sqrt{285,5\left(\frac{1}{120} - \frac{1}{200}\right)}} = +0,188$$

δ. Hệ số tương quan thứ tự (r_s)

Đối với các tình trạng có thể xếp thứ tự từ nhỏ đến lớn những đại lượng quan sát hoặc đo lường được, người ta có thể dùng phương pháp tính hệ số tương quan của Spiếc man (gọi là phương pháp tính hệ số tương quan thứ tự) để xác định hệ số tương quan. Phương pháp này tính tương đối nhanh, ít nhầm lẫn.

Công thức thực hành để tính là hệ số tương quan thứ tự r_s như sau:

$$r_s = 1 - \frac{6D^2}{n\left(n^2 - 1\right)} \qquad (11\text{-}39)$$

Trong đó: r_s là hệ số tương quan thứ tự

D là hiệu thứ tự của một cặp biến số

n là số cặp quan sát hoặc đo lường.

Muốn tính hệ số tương quan thứ tự r_s, trước tiên phải xếp thứ tự từ cao đến thấp các đại lượng quan sát hoặc đo lường được của các tính trạng. Trong khi xếp thứ tự các đại lượng đó mà gặp 2, ba... đại lượng bằng nhau thì thứ tự của các đại lượng này là trung bình cộng của hai, ba... số thứ tự kế tiếp số thứ tự của đại lượng đã được xác định thứ tự ở trên. Sau đó lập bảng tính hệ số tương quan thứ tự.

Thí dụ 1: Xác định hệ số tương quan giữa tỉ lệ mỡ sữa và tỉ lệ đạm sữa của bò theo tài liệu dưới đây (xem bảng 11-10).

Bảng 11-10. Tỉ lệ mỡ sữa và tỉ lệ đạm sữa của 10 bò

Số hiệu bò	1	2	3	4	5	6	7	8	9	10
Tỉ lệ mỡ sữa (%)	3,7	3,6	3,8	3,5	3,7	3,3	3,5	3,1	3,2	3,3
Tỉ lệ đạm sữa (%)	3,6	3,4	3,5	3,3	3,6	3,2	3,4	3,0	3,1	3,3

Xếp thứ tự từ cao đến thấp tỉ lệ mỡ sữa và tỉ lệ đạm sữa của 10 bò, lập bảng tính hệ số tương quan thứ tự (xem bảng 11-11).

Bảng 11-11. Bảng tính hệ số tương quan thứ tự

Số thứ tự (của bảng tính)	Tỉ lệ mỡ sữa (x)		Tỉ lệ đạm sữa (y)		D	D^2
	%	Thứ tự	%	Thứ tự		
1	3,8	1	3,5	3	-2	4
2	3,7	2,5	3,6	1,5	+1	1
3	3,7	2,5	3,6	1,5	+1	1
4	3,6	4,0	3,4	4,5	-0,5	0,25
5	3,5	5,5	3,4	4,5	+1	1
6	3,5	5,5	3,3	6,5	-1	1
7	3,3	7,5	3,3	6,5	+1	1
8	3,3	7,5	3,2	8	-1	1
9	3,2	9	3,1	9	0	0
10	3,1	10	3,0	10	0	0
Cộng						$\Sigma D^2 = 10,25$

Thay các giá trị của D^2 và n vào công thức 11-39 được:

$$r_s = 1 - \frac{6 \times 10,25}{10\left(10^2 - 1\right)} = +0,928$$

Thí dụ 2: Xác định hệ số tương quan giữa mức độ sử dụng lợn đực giống và thời gian sử dụng lợn đực giống theo tài liệu dưới đây: (xem bảng 11-12)

Bảng 11-12. Mức độ sử dụng và thời gian sử dụng của 7 lợn đực giống

Tên lợn đực giống	A	B	C	E	D	G	F
Thứ tự của mức độ sử dụng	1	3	2	5	4	7	6
Thứ tự của thời gian sử dụng	5	7	4	2	6	3	1

Xếp thứ tự nhiều đến ít về mức độ sử dụng và từ dài đến ngắn về thời gian sử dụng của 7 lợn đực giống, lập bảng tính hệ số tương quan thứ tự (xem bảng 11-13).

Bảng 11-13. Bảng tính hệ số tương quan thứ tự về mức độ sử dụng và thời gian sử dụng của 7 lợn đực giống

Số thứ tự (của bảng tính)	Mức độ sử dụng		Thời gian sử dụng		D	D^2
	Tên lợn đực giống	Thứ tự	Tên lợn đực giống	Thứ tự		
1	A	1	A	5	-4	16
2	B	2	B	4	-2	4
3	C	3	C	7	-4	16
4	D	4	D	6	-2	4
5	E	5	E	2	+3	9
6	F	6	F	1	+5	25
7	G	7	G	3	+4	16
Cộng						$\Sigma D_2 = 90$

Thay các giá trị của D^2 và n vào công thức 11-39 được:

$$r_s = 1 - \frac{6 \times 90}{7\left(7^2 - 1\right)} = -0,61$$

θ. Chỉ số liên hệ (ρ, đọc là rô)

Ngoài phương pháp xác định các hệ số tương quan đã giới thiệu ở trên, người ta còn dùng chỉ số liên hệ ρ để xác định mức độ tương quan, nhất là trong trường hợp cần xác định mức độ tương quan giữa một tính trạng định lượng và một tính trạng định tính.

Công thức thực hành để tính chỉ số liên hệ ρ như sau:

$$\rho = \frac{\alpha - 1}{\sqrt{\left(l_x - 1\right)\left(l_y - 1\right)}} \qquad (11\text{-}40)$$

Trong đó: $$\alpha = \sum\left[\frac{\sum\left(f_{xy}^2 : f_y\right)}{f_x}\right] - \frac{\left(l_x - 1\right)\left(l_y - 1\right)}{n} \qquad (11\text{-}41)$$

hoặc: $$\alpha = \sum\left[\frac{\sum\left(f_{xy}^2 : f_y\right)}{f_y}\right] - \frac{\left(l_x - 1\right)\left(l_y - 1\right)}{n} \qquad (11\text{-}42)$$

l_x : số tổ của tính trạng thứ nhất (x)

l_y : số tổ của tính trạng thứ hai (y)

f_{xy} : tần số đồng thời của tính trạng x và tính trạng y

f_x : tần số của các tổ theo tính trạng x

f_y : tần số của các tổ theo tính trạng y

n : dung lượng quan sát hoặc đo lường.

Để tính mức độ tương quan cần phải lập một bảng tính.

Thí dụ: Xác định mức độ tương quan giữa độ cong của lông và loại hình thể chất của cừu theo tài liệu dưới đây (xem bảng 11-14).

Bảng 11-14. Kết quả phân loại về độ cong của lông và loại hình thể chất của 100 cừu

Độ cong của lông / Loại hình thể chất	Lượn sóng	Vặn quả đậu	Uốn cong bờm ngựa	Xoắn vặn nút chai
Thanh	5	20	5	-
Thô	2	5	15	8
Chắc	15	20	5	-

Lập bảng tính mức độ quan hệ (xem bảng 11-15).

Bảng 11-15. Bảng tính mức độ quan hệ theo chỉ số liên hệ ρ

Loại hình thể chất (x)	Độ cong của lông (y)				f_1	$\sum \frac{f_{xy}^2}{f_y}$	$\sum \frac{f_{xy}^2}{f_y} : f_x$
	Lượn sóng	Vặn quả đậu	Uốn cong bờm ngựa	Xoắn vặn nút chai			
Thanh	5	20	5			1,136+8,888	
Thô	$5^2 = 25$	$20^2 = 400$	$5^2 = 25$	-	30	+1,0 = 11,024	$\frac{11,024}{30} = 0,367$
Chắc	$\frac{25}{22} = 1,136$	$\frac{400}{45} = 8,888$	$\frac{25}{25} = 1$				
	2	5	15	8		0,181+0,555	
	$2^2 = 4$	$5^2 = 25$	$15^2 = 225$	$8^2 = 64$	30	+ 9 + 8 = 17,736	$\frac{17,736}{30} = 0,591$
	$\frac{4}{22} = 1,181$	$\frac{25}{45} = 0,555$	$\frac{225}{25} = 9$	$\frac{64}{8} = 8$			
	15	20	5			10,227+8,888 + 1,0 = 20,115	$\frac{20,115}{40} = 0,503$
	$15^2 = 225$	$20^2 = 400$	$5^2 = 25$	-	40		
	$\frac{225}{22} = 10,227$	$\frac{400}{45} = 0,888$	$\frac{25}{25} = 1$				
f_2	22	45	25	8	n = 100		

Do đó: $$\sum\left[\frac{\sum\left(f_{xy}^2 : f_y\right)}{f_x}\right] = 0,367 + 0,591 + 0,503 = 1,461$$

Trong thí dụ này l_x là số tổ của loại hình thể chất và $l_x = 3$.

Còn l_y là số tổ của độ cong của lông và $l_y = 4$.

Từ đó ta có:

$$\alpha = 1,461 - \frac{(3-1)(4-1)}{100} = 1,461 - 0,06 = 1,401$$

Vậy chỉ số liên hệ ρ:

$$\rho = \frac{1,401 - 1}{\sqrt{(3-1)(4-1)}} = 0,164$$

Chỉ số liên hệ ρ luôn luôn là một số dương, vì thế nó không thể hiện hướng tương quan, mà nó chỉ thể hiện mức độ tương quan. Chỉ số này càng lớn thể hiện mức độ tương quan càng chặt chẽ, ngược lại chỉ số này càng nhỏ thể hiện mức độ tương quan càng ít chặt chẽ.

c. Kiểm tra mức độ tin cậy của hệ số tương quan

α. Phân bố tần số của các hệ số tương quan mẫu

Từ một tổng thể có hệ số tương quan ρ chọn ra nhiều mẫu, mỗi mẫu có một hệ số tương quan r, do tính chất ngẫu nhiên trong quá trình chọn mẫu nên các hệ số tương quan của mẫu đều khác nhau, mức độ khác nhau giữa các hệ số tương quan mẫu thể hiện ở phân bố tần số của các hệ số tương quan mẫu này. Phân bố tần số của các hệ số tương quan phụ thuộc vào hai nhân tố sau:

- Giá trị tuyệt đối của hệ số tương quan tổng thể ρ

Khi ρ = 0, các giá trị của r phân bố quanh O, nhỏ nhất là -1, lớn nhất là +1, do đó phân bố tần số của r là đối xứng qua đường thẳng trực quan với trục hoành ở điểm O (xem hình 11-17).

Khi ρ ≠ 0, thí dụ bằng 0,8, các giá trị của r phân bố quanh 0,8, nhỏ nhất vẫn là -1, lớn nhất vẫn là +1, nhưng khi r lớn hơn 0,8 thì các giá trị của r thay đổi từ 0,8 đến +1 (phạm vi thay đổi là 0,2) và khi r nhỏ hơn 0,8 thì giá trị của r thay đổi từ -1 đến 0,8 (phạm vi thay đổi là 1,8) nghĩa là phạm vi thay đổi của r quanh 0,8 không giống nhau, do đó phân bố tần số của r là không đối xứng qua đường thẳng trực giao với trục hoành điểm 0,8 (xem hình 11-17).

- Số lượng cặp quan sát hoặc đo lường

Nếu ρ ≠ 0, mức độ không đối xứng của phân bố hệ số tương quan mẫu r sẽ giảm dần khi số lượng cặp quan sát hoặc đo lường tăng lên; vì rằng khi dung lượng mẫu nhỏ thì