TRƯỜNG ĐẠI HỌC NÔNG NGHIỆP I HÀ NỘI
KHOA CHĂN NUÔI - THÚ Y

BÀI GIẢNG
PHƯƠNG PHÁP THÍ NGHIỆM TRONG CHĂN NUÔI & THÚ Y

(PHẦN I)

Đỗ Đức Lực

Bộ môn Di truyền - Giống, Khoa Chăn nuôi - Thú y

Hà Nội - 2004

MỤC LỤC

Bài giảng môn học *Phương pháp thí nghiệm trong chăn nuôi thú y* được soạn riêng cho sinh viên chuyên ngành chăn nuôi & thú y, hệ chính quy. Bài giảng này bao gồm 2 phần; đây là phần I, bao gồm 2 chủ đề chính là *Tóm tắt dữ liệu và Ước lượng & Kiểm định giả thiết*; phần II sẽ được in riêng với 2 chủ đề chính là *Bố trí thí nghiệm và tương quan & hồi quy*.

Mặc dù có rất nhiều cố gắng trong quá trình biên soạn, xong không thể tránh được những thiếu sót. Tác giả rất mong sự góp ý của bạn đọc. Mọi ý kiến góp ý xin gửi theo địa chỉ sau đây:

Đỗ Đức Lực

Phòng 303 & 304

Bộ môn Di truyền - Giống, Khoa Chăn nuôi - Thú y

Đại học Nông nghiệp I Hà Nội, Trâu Quỳ, Gia Lâm

E-mail: dtghn@yahoo.co.uk

Điện thoại Bộ môn: 04 - 876 82 65

Giới thiệu chung

Trong khoá học *Phương pháp thí nghiệm trong chăn nuôi và thú y* sẽ đề cập đến 4 nội dung chính sau đây:

- Tóm tắt và mô tả số liệu
- Ước lượng và Kiểm định giả thuyết
- Các nguyên tắc cơ bản và một số mô hình thiết kế thí nghiệm thường gặp trong chăn nuôi và thú y.
- Tương quan và hồi quy.

Khoá học sẽ cung cấp cho sinh viên chuyên ngành chăn nuôi thú y nắm được cách phân tích số liệu, các nguyên tắc bố trí một thí nghiệm và rút ra những kết luận từ việc phân tích số liệu.

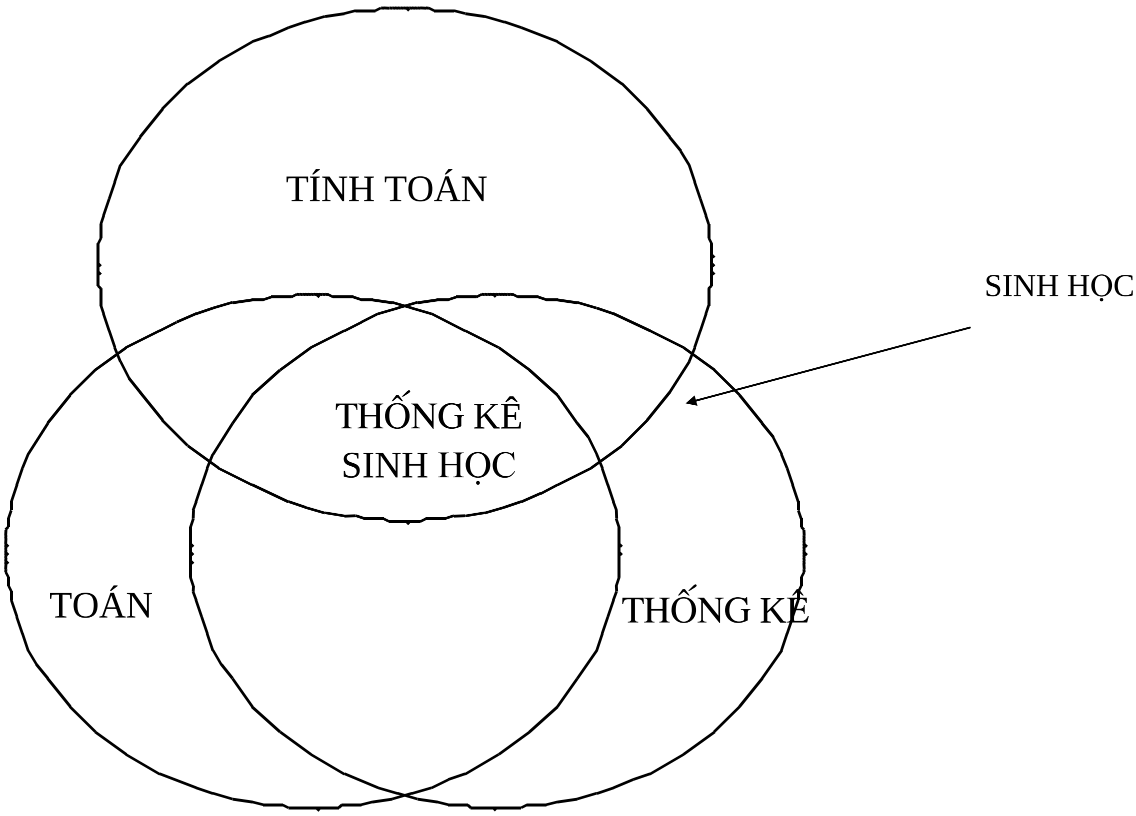
Tổng số thời lượng của khoá học là 2 đơn vị học trình (30 tiết), trong đó phần lý thuyết 20 tiết và thực hành 10 tiết. Các bài thực hành được thực hiện tại Phòng máy tính Khoa Chăn nuôi - Thú y (Phòng 218, tầng 2).

Trong suốt khoá học sẽ có 5 bài kiểm tra; điểm số của mỗi bài kiểm tra được nhân với hệ số 0,1 nhưng chỉ lấy 4 bài có điểm số cao nhất để tính vào điểm cuối kỳ. Kết thúc khoá học sẽ có một bài thi cuối kỳ; điểm số của bài thi được nhân với hệ số 0,6. Điểm đánh giá của môn học chính là tổng số điểm của 4 bài kiểm tra và bài thi cuối kỳ sau khi đã nhân với các hệ số tương ứng. Học viên được sử dụng tài liệu trong quá trình làm bài kiểm tra hoặc bài thi.

1. Khái niệm về các biến sinh học

1.1. Các vấn đề sẽ đề cập tới

- Thống kê sinh học là gì?
- Các kiểu biến trong sinh học
- Các ví dụ minh hoạ



1.2. Thống kê sinh học là gì?

Nếu hiểu một cách chính xác, thống kê sinh học có nghĩa là chắc nghiệm trong sinh học. Một định nghĩa hiện đại và tổng quát hơn là: Sử dụng thống kê, toán học và các phương pháp tính toán để trả lời các câu hỏi về sinh học.

Trong suốt khoá học chúng ta sẽ tập chung vào hai vấn đề có liên quan mật thiết trong thống kê sinh học: **phương pháp thiết kế thí nghiệm** và **phân tích thống kê** các số liệu được thu thập từ các mô hình định trước. Những kỹ thuật phân tích được sử dụng đối với các số liệu thu thập từ các thí nghiệm được bố trí cũng được áp dụng đối với các số liệu từ các **nghiên cứu quan sát**. Thiết kế thí nghiệm đóng một vai trò quan trọng và thường được sử dụng trong thú y.

1.3. Các dạng biến trong sinh học

1.3.1. Giới thiệu

Các nghiên cứu trong chăn nuôi thú y, chúng ta phải thường xuyên làm việc với các dữ liệu. Các dữ liệu có thể bằng số, bằng chữ hặc các ký hiệu…, chúng được đặc trưng cho một cá thể, một một nhóm hay một quần thể. Các dữ liệu như vậy ta thường gọi là *biến sinh học* hay thường gọi tắt là *biến*.

Nếu ta tiến hành các thí nghiệm sinh học nhiều lần được đặt dưới cùng một điều kiện, số liệu thu được trong mỗi lần quan sát đều khác nhau bởi có *sự biến động sinh học tự nhiên.* Sự biến động này do yếu tố di truyền và yếu tố môi trường tác động lên.

Ví dụ điển hình, năng suất sữa của bò sữa tăng không như nhau trong cùng một điều kiện. Nó biến động từ ngày này qua ngày khác và giữa các con bò cũng khác nhau. Đây chính là sự khác biệt giữa các ngành khoa học sinh học với các ngành khoa học khác như vật lý hay hoá học. Nếu một quả bóng được ném từ độ cao xác định thì thời gian từ khi quả bóng rơi đến khi chạm đất coi như gần bằng nhau. Nếu thực hiện phản ứng hoá học xác định thì khối lượng sản phẩm tạo ra từ phản ứng hoá học là như nhau đối với mỗi lần.

Số liệu trong sinh học thì hoàn toàn khác xa do ảnh hưởng tương tác giữa kiểu gen và môi trường. Số liệu thu được cũng có thể rất khác nhau bởi vì trong thực tế chúng ta không thể lặp lại thí nghiệm dưới cùng một điều kiện. Vì vậy để kiểm soát được sự biến động này, thiết kế thí nghiệm đóng vai trò rất quan trọng trong nghiên cứu.

1.3.2. Phân loại biến

Chúng ta có thể phân loại các dạng biến, mà có thể thường gặp như sau:

Biến

Biến định lượng — Biến định tính

Biến liên tục — Biến rời dạc — Biến thứ hạng — Biến thuộc tính

1.3.2.1. Biến định lượng

Các giá trị có thể thể hiện được và đo đạc được dưới dạng số. Trong sinh học chúng có thể được xem xét như các "*tính trạng số lượng*".

Biến liên tục: biến có thể (về lý thuyết) có giá trị không hạn chế, thậm chí nằm ở vùng giới hạn.

Ví dụ: Trọng lượng cơ thể (kg); tỷ lệ nạc (%), chiều cao (cm)...

Biến rời dạc: các giá trị được giới hạn trong khoảng nhất định (không có những điểm trung gian). Thông thường biến rời dạc là những giá trị đếm được (giá trị 0, 1, 2, 3,...)

Ví dụ: Số con sinh ra trong một lứa, tế bào bạch cầu đếm được trên kính hiển vi.

1.3.2.2. Biến định tính

Các giá trị không thể biểu diễn được bằng số thực nhưng có thể xếp hạng được. Chúng được gọi là các "*tính trạng chất lượng*".

Biến thứ hạng: Các giá trị định tính có thể thay thế theo một thứ tự có ý nghĩa nào đó.

Ví dụ: mức độ dễ đẻ của bò (1 = "đẻ thường", 2 = "đòi hỏi sự can thiệp ở một số khâu", 3 = "đòi hỏi sự can thiệp của các bác sỹ thú y"); mức độ nhiễm bệnh , đối với trường hợp này, mỗi một mức độ bệnh được ấn định bằng một số (0 = "không nhiễm bệnh", 1 = "nhiễm bệnh"

Biến thuộc tính: Các giá trị định tính không thể sắp xếp theo một thứ tự nào cả.

Ví dụ: Kiểu gen (đồng hợp tử, dị hợp tử...), dạng tế bào máu (basophils, eosinophils, lymphocytes...), các giống vật nuôi khác nhau.

1.4. Bài tập:

Dựa vào phân loại của các biến sinh học, anh (chị) lấy ít nhất 2 ví dụ trong chuyên ngành chăn nuôi thú y cho từng loại biến. Để thực hiện được bài tập các anh (chị) có thể tìm các bài báo khoa học, các báo cáo tốt nghiệp đại học, các luận văn thạc sỹ, tiến sỹ... để từ các thí nghiệm trong đã được bố trí; xác định xem các biến đã nghiên cứu thuộc nhóm nào.

Lưu ý: Có thể tham khảo *Tạp chí Khoa học Nông nghiệp* của ĐH Nông nghiệp I trực tuyến theo địa chỉ website sau: http://www.hau1.edu.vn/tapchi_KHNN.htm

2. Tóm tắt và trình bày các dữ liệu

2.1. Các vấn đề sẽ đề cập tới

- Tóm tắt dữ liệu
- Biểu đồ và tổ chức đồ
- Tổng thể và mẫu
- Các tham số thống kê mô tả

2.2. Giới thiệu

Bản thân số liệu thô không nói lên ý nghĩa gì. Nó chỉ thực sự có giá trị khi ta có thể rút ra những kết luận từ số liệu đó. Để có thể rút ra những thông tin tóm tắt hữu ích từ số liệu thô thì chúng ta cần phải thay thế số liệu thô bằng số liệu tinh *dưới dạng số* hoặc *đồ thị*. Tóm tắt dữ liệu bao gồm các thông tin về phân phối số lượng phân phối tần suất, các tham số chỉ vị trí (trung bình, trung vị, mode) và mức độ phân tán (phương sai, biên độ dao động, hệ số biến động).

2.3. Phân phối tần suất

2.3.1. Phân phối tần suất của các tính trạng chất lượng

Khi dữ liệu thu được dưới dạng thứ hạng hoặc thuộc tính (biến định tính), mỗi một quan sát sẽ trở thành các nhóm hoặc thứ hạng. Chúng ta có thể dùng biểu đồ dạng cột hoặc dạng bánh để biểu diễn số hoặc phần trăm của từng nhóm.

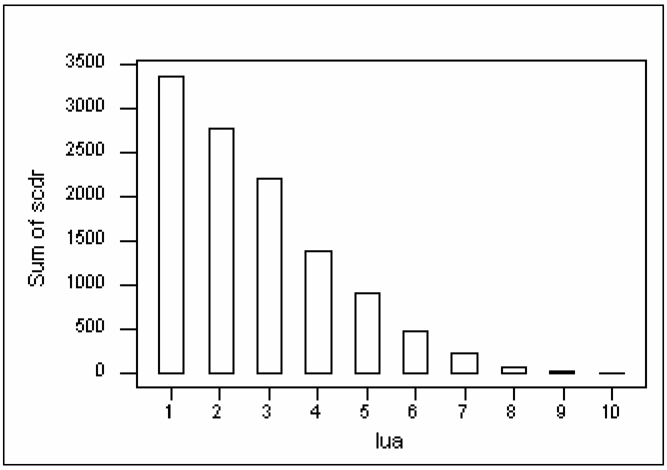
Ví dụ: Số con đẻ ra qua các lứa được theo dõi tại trại Mỹ Văn từ năm 1996 đến năm 2001 (số liệu được lấy từ đề tài cấp Nhà nước):

Lứa	Số con đẻ ra (con)	Tần suất (%)	Tần suất tích luỹ (%)
1	337	29.82	29.82
2	275	24.34	54.16
3	213	18.85	73.01
4	137	12.12	85.13
5	86	7.61	92.74
6	49	4.34	97.08
7	22	1.95	99.03
8	8	0.71	99.73
9	2	0.18	99.91
10	1	0.09	100.00

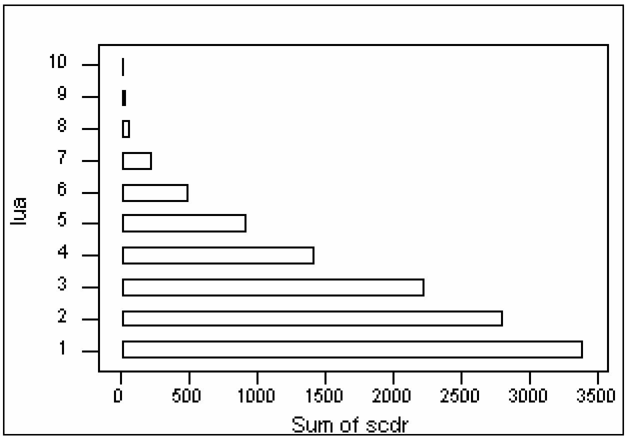
- Biểu đồ dạng cột

Trong biểu đồ dạng cột từng nhóm trong một biến được thể hiện dưới dạng cột. Diện tích của các cột và các khoảng trống ở trục hoành đều không có ý nghĩa; điều quan trọng nhất là chiều cao (nếu là cột thẳng đứng) hoặc chiều dài (nếu là cột nằm ngang) của các cột. Chiều cao hoặc chiều rộng sẽ tỷ lệ với phần trăm của từng nhóm.

Ví dụ: Biểu đồ về số con đẻ ra qua các lứa tại trại Mỹ Văn từ năm 1996 đến 2001



Biểu đồ dạng cột đứng

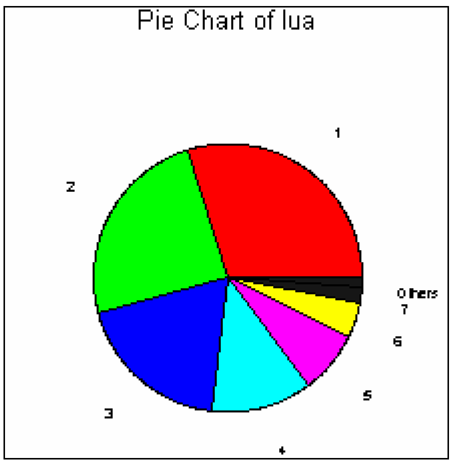


Biểu đồ dạng cột nằm ngang

- Biểu đồ dạng bánh

Biểu đồ dạng bánh hình tròn dùng để biểu diễn dữ liệu thuộc các lớp hoặc các nhóm khác nhau bằng các miếng tỷ lệ với tần suất hoặc số lượng tương ứng. Biểu đồ dạng bánh cũng thường được sử dụng để so sánh, vì tỷ lệ dưới dạng miếng dễ quan sát hơn bằng mắt thường hơn là chiều cao của từng cột.

Ví dụ: Biểu đồ dạng bánh về số con đẻ ra qua các lứa



2.3.2. *Phân phối tần suất của các tính trạng số lượng (dữ liệu 1 chiều)*

Ta sử dụng tổ chức đồ và đồ thị để biểu diễn các dữ liệu định lượng.

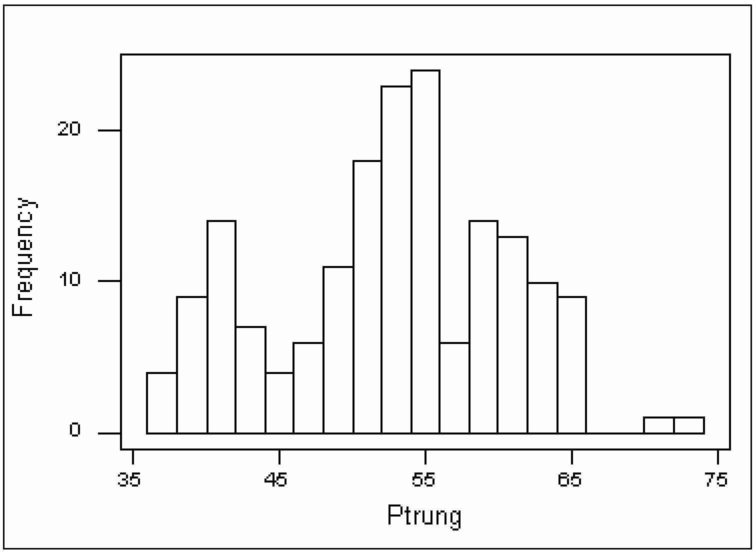
- Tổ chức đồ

Phân bố tần suất hoặc số lượng của biến liên tục có thể biểu diễn dưới dạng tổ chức đồ. Trong tổ chức đồ diện tích của từng hình chữ nhật tỷ lệ với tần suất hoặc số lượng trong từng khoảng.

Ví dụ: Khối lượng (g) của 174 quả trứng gà cân được tại trại Quang Trung, Trường ĐH Nông nghiệp I Hà Nội (số liệu được lấy từ đề tài nhóm sinh viên nghiên cứu khoa học năm học 2002 - 2003)

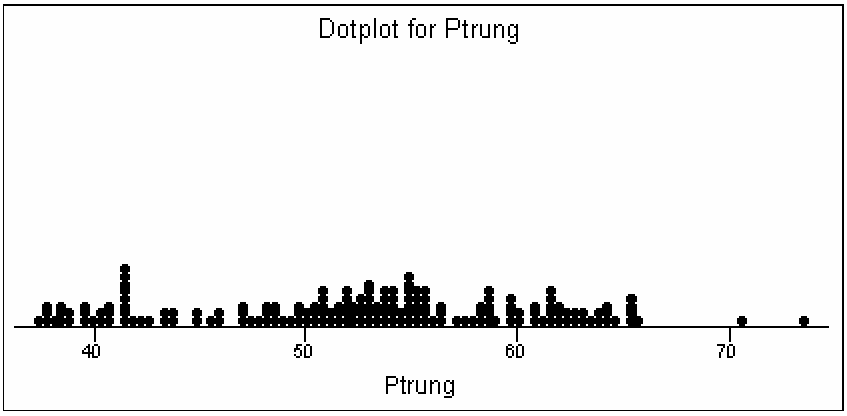
54,9	54,0	55,8	50,4	55,3	50,3	53,1	50,9	50,9	53,8
54,5	52,2	54,3	55,5	51,8	53,6	52,5	48,5	52,8	55,0
52,3	52,0	52,0	53,1	55,8	53,4	51,2	49,5	52,6	54,7
56,4	56,1	55,4	53,5	44,7	64,4	55,4	54,8	55,5	58,7
65,6	59,9	65,5	48,0	65,5	55,0	55,0	55,0	62,2	61,6
46,1	50,0	53,5	53,0	61,5	62,0	61,1	58,6	59,7	52,6
50,6	54,2	63,1	53,6	61,0	58,2	53,9	50,6	55,5	57,5
65,2	61,0	61,6	63,0	58,0	58,6	58,4	58,7	65,2	61,8
60,7	63,7	62,2	63,4	64,1	63,7	73,4	62,7	61,5	59,9
58,2	54,2	53,8	49,4	60,3	64,6	61,5	59,0	70,4	61,8
64,2	59,8	56,2	62,9	56,5	37,9	43,3	39,4	41,3	41,3
41,6	43,8	39,4	42,3	40,8	40,0	41,3	37,9	45,8	41,4
40,6	40,4	45,4	38,4	37,5	42,0	38,6	37,8	40,3	41,3
38,5	43,3	42,6	38,2	43,7	41,6	38,8	39,0	39,4	51,7
49,7	51,7	50,7	47,6	54,8	52,9	52,9	54,0	41,6	50,3
52,1	47,9	49,1	47,0	49,8	51,9	48,6	48,6	60,0	52,9

Ta biểu diễn tần suất của 174 quả trứng này bằng tổ chức đồ sau



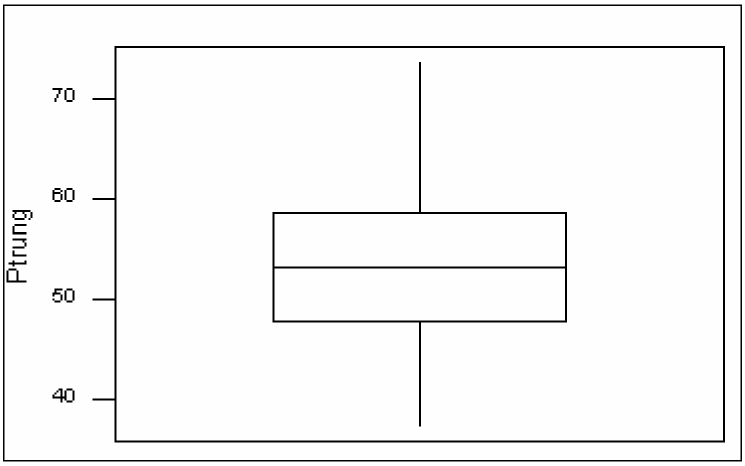
- Đồ thị điểm

Nếu số liệu quan sát ở mức độ giới hạn, thì tốt nhất ta biểu diễn từng quan sát dưới dạng đồ thị điểm.



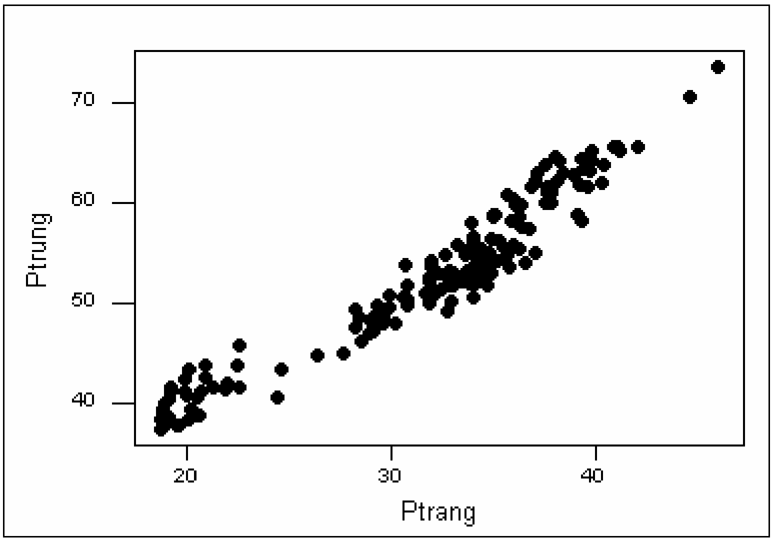
- Đồ thị hộp

Một số chương trình máy tính cho ta một dạng đồ thị mới kiểu như một cái hộp, vì vậy chúng ta gọi là đồ thị dạng hộp. Kiểu đồ thị này được sử dụng để mô tả dữ liệu của biến liên tục



2.3.3. *Tóm tắt và biểu diễn dữ liệu các tính trạng số lượng (dữ liệu 2 chiều)*

Đồ thị phân tán được sử dụng một cách rất hữu hiệu khi ta quan tâm đến mối liên hệ giữa 2 biến liên tục. Đồ thị được xây dựng khi ta vẽ n các điểm trên hệ toạ độ, các điểm này có toạ độ là x_iy_i. Đồ thị sau đây biểu diễn mối liên hệ giữa khối lượng quả trứng gà với khối lượng lòng trắng trứng của 174 quả (đề tài nghiên cứu của sinh viên lớp CN45A năm học 2002 - 2003).



2.4. Các số đo về vị trí và mức độ phân tán

2.4.1. Mẫu và tổng thể

2.4.1.3. Tổng thể

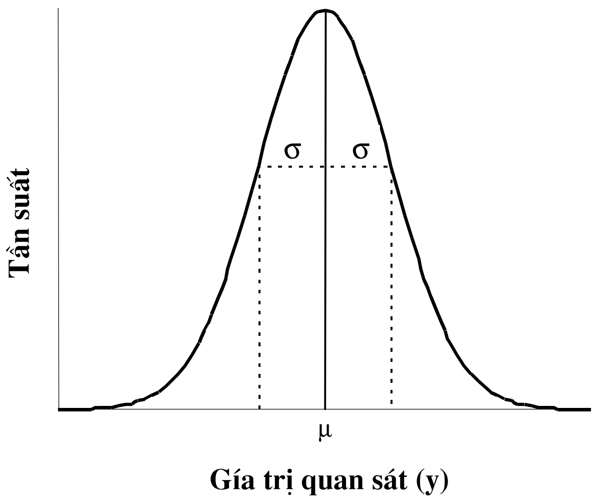
Tổng thể là tập hợp tất cả các thành viên có cùng một đặc tính nhất định. Tổng thể có thể là có thực và chính vì vậy có thể liệt kê ra, ví dụ số lượng lợn nái ở các trại lợn giống ở các tỉnh phía Bắc Việt Nam. Chúng cũng có thể chỉ giả thiết và không thể liệt kê được, ví dụ số lợn nái hiện có ở Việt Nam.

Đặc trưng của tổng thể là rất lớn - thậm chí là không hạn chế! Tổng thể có thể được miêu tả bằng những ***tham số của tổng thể*** (ký hiệu bằng các chữ cái Hy Lạp)

Trung bình quần thể = μ

Phương sai quần thể = σ^2

Trong suốt khoá học này, ta luôn giả sử rằng phân phối tần suất của quần thể nghiên cứu luôn có phân bố chuẩn với trung bình quần thể = μ, và phương sai quần thể = σ^2.



Dạng rút gọn: $y \sim N(\mu,\sigma^2)$. Đọc là: Biến y có phân bố chuẩn với trung bình μ và phương sai σ^2

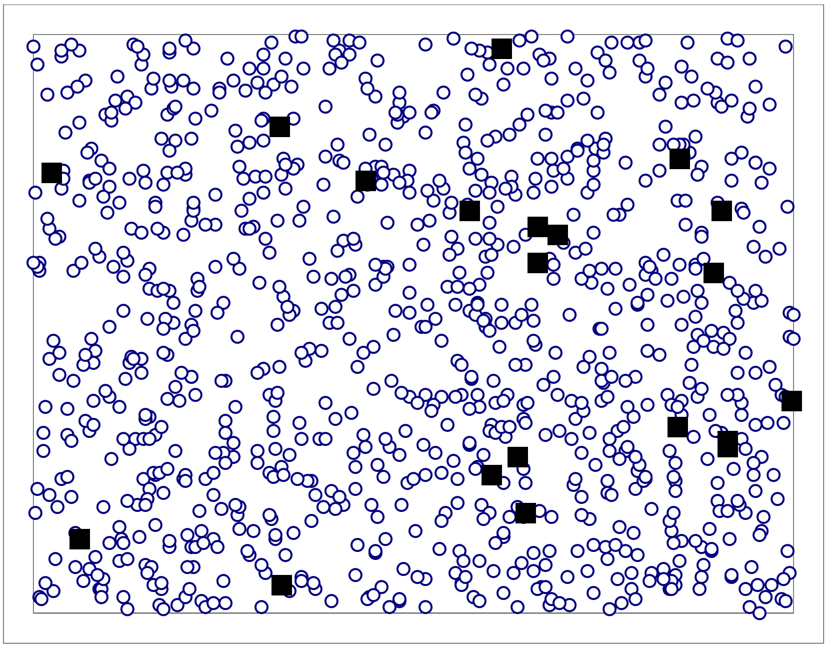
Đối với phân bố chuẩn ta luôn có:

68% số quan sát nằm trong khoảng $\mu \pm 1\sigma$
95% số quan sát nằm trong khoảng $\mu \pm 2\sigma$
99,7% số quan sát nằm trong khoảng $\mu \pm 3\sigma$

Từ một quần thể lớn, chúng ta thường khó xác định các giá trị này một cách chính xác. Nếu ta tiến hành nghiên cứu toàn bộ các cá thể của một quần thể. Công việc này đòi hỏi rất nhiều thời gian và kinh phí; nếu đứng trên phương diện kinh tế thì không hiệu quả. Tiến hành nghiên cứu một tổng thể đôi khi cho ta kết quả không chính xác; do có nhiều người tham gia và cũng có rất nhiều phương tiện đo đạc khác nhau ở những thời điểm khác nhau dẫn đến sai số rất lớn. Xuất phát từ thực tế trên, trong nghiên cứu chỉ tập trung nghiên cứu trên các ***mẫu đại diện***.

2.4.1.4. Mẫu

Chúng ta có thể chọn một mẫu (dung lượng mẫu n) từ quần thể một cách "ngẫu nhiên". Ví dụ: n = 20 mẫu (■) được chọn một cách ngẫy nhiên từ một quần thể N = 1,000 (○)



Mẫu được chọn một cách đại diện cho quần thể - nhưng cách chọn này ***không có gì đảm bảo*** là đã chọn được một mẫu đại diện. Vì vậy để kết quả có tin cậy cao cần phải có sự lặp lại trong việc rút mẫu nghiên cứu.

Nghiên cứu trên các mẫu đại diện sẽ dễ dàng hơn, nhanh hơn và rẻ hơn so với việc nghiên cứu cả quần thể ($n << N$).

Giá trị trung bình của mẫu nghiên cứu được ký hiệu bằng các chữ cái có dấu ngang ở phía trên, ví dụ như $\bar{x}$, $\bar{y}$ hoặc với các chỉ số dưới như $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$...

Từ các số đo của mẫu ta có thể sử dụng các giá trị đó để ước tính cho quần thể:

Trung bình mẫu $(\bar{y})$ $\rightarrow$ Trung bình quần thể (μ)

Phương sai mẫu (s^2) $\rightarrow$ phương sai quần thể (σ^2)

- Lưu ý

Nếu 1 biến x có phân bố với trung bình μ và σ^2, thì biến $\bar{x}$ là giá trị trung bình của mẫu với *n* quan sát của biến *x* sẽ có phân bố với trung bình μ và phương sai σ^2/n

2.4.2. Các các số đo về vị trí và mức độ phân tán

2.4.2.5. Các vấn đề sẽ đề cập tới

- Các số đo về vị trí
  - Trung bình
  - Trung vị
  - Mode
- Các số đo về mức độ phân tán
  - Phương sai
  - Độ lệch chuẩn
  - Miền tứ vị

Ví dụ: Mead và cộng sự (1993) trang34

Ba trại sử dụng các phương pháp chăn nuôi lợn khác nhau. Sử dụng các giống lợn tương tự nhau. Thời gian từ lúc cai sữa đến xuất bán được ghi lại như sau (ngày):

Trại 1	Trại 2	Trại 3
105	107	100
112	108	107
99	104	100
97	112	113
104	101	103
117	103	115
	105	98
	108	110
		105

2.4.2.6. Các tham số chỉ vị trí

Trung bình cộng

- Công thức tính:

$$\boxed{\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\left(y_1 + y_2 + \ldots + y_n\right)}$$

- Ví dụ (số liệu ở trại thứ 3)

$$\bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = \frac{1}{9}\left(y_1 + y_2 + y_3 + \ldots + y_9\right)$$
$$= \frac{1}{9}(100 + 107 + \ldots + 105) = 105{,}7$$

Trung vị (Median)

- Trung vị được ký hiệu là *M*

Là giá trị nằm chính giữa bộ số liệu: 50% số quan sát ở phía dưới trung vị và 50% ở trên. Lợi ích của trung vị là khi dữ liệu chứa các giá trị rất lớn với tần số thấp chúng sẽ ảnh hưởng mạnh đến trung bình số học, trong khi đó chúng hầu như không ảnh hưởng đến giá trị trung vị. Do đó lúc này trung vị cho ta một ý niệm tốt hơn về giá trị trung tâm của phân phối.

- Công thức tính

Trước hết ta sắp xếp số liệu theo thứ tự tăng dần

Đánh số thứ tự cho các dữ liệu sau khi đã sắp xếp theo thứ tự tăng dần

Tìm trung vị theo công thức với dung lượng mẫu là n, *M* = (n+1) / 2

Lưu ý rằng trong công thức nêu trên n không phải là dung lượng mẫu trong thí nghiệm mà là số thứ tự lớn nhất sau khi đã được đánh số.

- Ví dụ (đối với trại thứ nhất)

Sắp xếp số liệu theo thứ tự tăng dần và đánh số thứ tự

98	100	100	103	105	107	110	113	115
1	2	3	4	5	6	7	8	9

Trung vị

M = (n+1) / 2 = (9+1) / 2 =5; tức là trung vị nằm ở vị trí quan sát thứ 5 trong bảng số liệu đã sắp xếp thứ tự , tức là trung vị = $\tilde{y}$ = 105 ngày

Chú ý trung bình có giá trị tương tự (105.7 ngày)

- Ví dụ (đối với trại 2)

101	103	104	105	107	108	108	112
1	2	3	4	5	6	7	8

Trung vị

. Trung vị $=\left(\frac{8+1}{2}\right)$ = 4,5 giá trị đã sắp xếp theo thứ tự, tức là trung vị nằm giữa giá trị thứ 4 và thứ 5, hay trung vị là ½(105 + 107) = 106 ngày.

Mode

Là giá trị có tần suất cao nhất trong bộ dữ liệu. Trong phân bố tần suất, Mode là giá trị nằm ở điểm cao nhất trên đường cong. Đối với phân bố chuẩn thì Mode cũng chính là trung vị và trung bình.

Các tham số chỉ sự biến động

Bước tiếp theo chúng ta cần xác định mức độ biến động xung quanh các giá trị đặc trưng như độ lệch chuẩn hoặc phương sai, miền hoặc miền tứ vị.

Phương sai

Phương sai của quần thể được ký hiệu l à σ^2

Phương sai của mẫu được ký hiệu là s^2

- Công thức

Dưới dạng tổng quát, ta có ***n quan sát***, thì công thức tổng quát tính phương sai là

$$\boxed{s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

Đơn vị tính của phương sai luôn là ***đơn vị tính của quan sát bình phương***. Nếu đơn vị tính của phép đo là kg (ví dụ trọng lượng cơ thể), thì phương sai có đơn vị tính là kg^2

- Ví dụ (đối với trại thứ 3)

Trong trại thứ 3 ta có tất cả 9 quan sát, tức n = 9.

Phương sai = s^2

$$= \frac{1}{9-1}[(100-105.7)^2 + (107-105.7)^2 + ... + (105-105.7)^2]$$

$$= 36.5 \text{ ngày}^2$$

Độ lệch chuẩn

Độ lệch chuẩn của quần thể được ký hiệu l à σ

Độ lệch chuẩn của mẫu được ký hiệu là s

Để đơn vị đo mức độ biến động của có cùng đơn vị tính như đơn vị đo của các quan sát, ta tiến hành lấy căn bậc 2 của phương sai. Đây chính là ***độ lệch chuẩn*** của các quan sát (thường được ký hiệu là s).

- Công thức tính độ lệch chuẩn

$$\boxed{s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

- Ví dụ (đối với trại thứ 3)

$$s = \sqrt{s^2} = \sqrt{36,5} = 6,04 \text{ ngày}$$

Hệ số biến động (Cv)

Như chúng ta đã biết độ lệch chuẩn được dùng để xác định mức độ biến động của một quần thể. Nhưng một vấn đề đặt là từ độ lệch chuẩn ta có thể biết được biến động của quần thể A nhỏ hay lớn hơn quần thể B; khi giá trị trung bình của các quần thể so sánh khác nhau thì việc sử dụng phương sai hay độ lệch chuẩn để so sánh độ biến động, đặc biệt khi rút mẫu nghiên cứu qua chênh lệch nhau. Để khắc phục những hạn chế nêu trên, chúng ta sử dụng một tham số thống kê ***hệ số biến động***.

- Công thức

$$\boxed{Cv = \frac{s \times 100}{\bar{y}}}$$

- Ví dụ (đối với trại thứ 3)

ta có: $\bar{y} = 105,7$ ngày và $\bar{s} = 6,04$ ngày $\rightarrow Cv = \dfrac{s \times 100}{\bar{y}} = \dfrac{6,04 \times 100}{105,7} = 5,74\,\%$

Sai số tiêu chuẩn (độ lệch chuẩn của giá trị trung bình)

Đối với các giá trị trung bình, người ta sử dụng sai số tiêu chuẩn của giá trị trung bình thay thế cho S.

Công thức

$$\boxed{S_{\overline{X}} = \frac{S}{\sqrt{n}}}$$

Ví dụ (đối với trại thứ 3)

ta có: $\bar{s} = 6,04$ ngày và n = 9 $\rightarrow S_{\overline{X}} = \dfrac{S}{\sqrt{n}} = \dfrac{6,04}{\sqrt{9}} = 2,01$

Miền tứ vị (IQR)

Thông thường để miêu tả sự biến động xung quanh giá trị trung bình, chúng ta xác định số lượng quan sát trong một miền như chia trung vị của mẫu cho 2, toàn miền chia thành 4 nhóm:

25% quan sát ≤ miền tứ vị dưới (Q_1)

50% quan sát ≤ trung vị (Q_2)

75% quan sát ≤ miền tứ vị trên (Q_3)

Công thức

Tứ vị dưới = $Q_1 = \left(\dfrac{n+1}{4}\right)$ giá trị đã được xếp hạng

Tứ vị trên = $Q_3 = \left(\dfrac{3(n+1)}{4}\right)$ giá trị đã được xếp hạng

Dạng tổng quát tính mức phần trăm thứ X = (n+1) X/100.

Ví dụ (đối với trại thứ 3) với số liệu đã đã được sắp xếp:

98	100	100	103	105	107	110	113	115

Tứ vị dưới　　Trung vị　　Tứ vị trên

Tứ vị dưới $= \left(\dfrac{n+1}{4}\right)$ giá trị đã được xếp hạng

$= \left(\dfrac{9+1}{4}\right)$ giá trị đã được xếp hạng

= 2.5 giá trị đã được xếp hạng

= tăng trọng trung bình giữa giá trị thứ 2 và thứ 3

= 0,5 × 100 + 0,5 × 110 = 100 ngày

Tứ vị trên $= \left(\dfrac{3(n+1)}{4}\right)$ giá trị đã được xếp hạng

$= \left(\dfrac{3(9+1)}{4}\right)$ giá trị đã được xếp hạng

= 7,5 giá trị đã được xếp hạng

= tăng trọng trung bình giữa giá trị thứ 7 và thứ 8

= 0,5 × 110 + 0,5 × 113 = 111,5 ngày

Như vậy Tứ vị dưới (Q_1) = 100 ngày

Tứ vị trên (Q_3) = 111,5 ngày

Với mức phần tử nhỏ hơn 30% ta có

= (n+1)X/100 = (9+1)30/100 = 3, giá trị này sẽ là 100 ngày.

Ta có khoảng cách giữa tứ vị trên và tứ vị dưới (IQR)

= Q3 - Q1 = 111,5 - 100 = 11,5

Những giá này thường bộc lộ cho ta nhiều thông tin hơn là các tóm tắt bằng số, như các tham số chỉ vị trí và biến động biểu hiện

Các giá trị min, max, Q1, Q2, Q2 và IQR được sử dụng để xác định những giá trị ngoại lai và trong một số trường hợp kiểm tra phân bố của số liệu.

Như ở ví dụ trên ta có các giá trị tương ứng là 98; 115; 100; 106; 111,5

Ta có 1,5×IQR = 1,5×11,5 = 17,25;

Như vậy giới hạn trên sẽ là Q3 + 1,5×IQR = 111,5 + 17,25 = 128,75 giới hạn dưới sẽ là Q1 - 1,5×IQR = 100 - 17,25 = 82,75

Với sự trợ giúp của các phần mềm thống kê ta có thể dễ dàng tóm tắt các dữ liệu một cách nhanh chóng và chính xác. Với ví dụ đã nêu trên, bằng phần mềm Excel hoặc Minitab ta có thể tính được các tham số thống kê mô tả như sau:

```
Descriptive Statistics: trai1, trai2, trai3


Variable            N       Mean     Median     TrMean      StDev    SE Mean
trail               6     105.67     104.50     105.67       7.63       3.12
trai2               8     106.00     106.00     106.00       3.46       1.22
trai3               9     105.67     105.00     105.67       6.04       2.01

Variable      Minimum    Maximum         Q1         Q3
trail           97.00     117.00      98.50     113.25
trai2          101.00     112.00     103.25     108.00
trai3           98.00     115.00     100.00     111.50
```

2.5. Bài tập

Khối lượng của 20 quả trứng (g) được trình bày dưới đây:

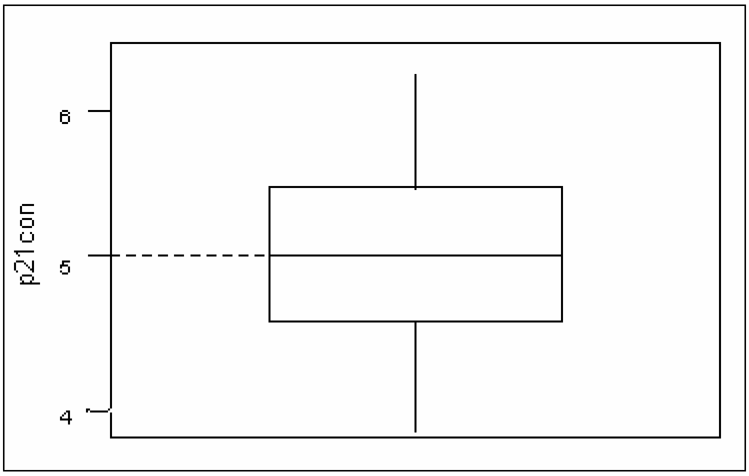
54,9	54,0	55,8	50,4	55,3	50,3	53,1	50,9	50,9	53,8
54,5	52,2	54,3	55,5	51,8	53,6	52,5	48,5	52,8	55,0

Hãy tính các tham số sau (bao gồm các ký hiệu và đơn vị đo tương ứng)

Tham số	Ký hiệu	Giá trị	Đơn vị tính
Trung bình			
Trung vị			
Mode			
Độ lệch chuẩn			
Phương sai			
Sai số tiêu chuẩn			
Hệ số biến động			

2.6. Bài kiểm tra số 1

Trong một thí nghiệm, 5 con lợn 21 ngày tuổi được rút một cách ngẫu nhiên từ một quần thể có khối lượng trung bình là 5,26 kg và độ lệch chuẩn là 0,65 kg. Sau khi mô tả khối lượng 21 ngày tuổi của 5 lợn nói trên bằng phần mềm *Minitab* ta thấy độ lệch chuẩn của mẫu bằng độ lệch chuẩn của quần thể và thu được đồ thị hộp:



1. (*2 điểm*) Anh (chị) hãy tóm tắt các tham số của đề ra bằng các ký hiệu thích hợp cùng với các đơn vị đo tương ứng

2. (*3 điểm*) Trong quần thể nói trên, có bao nhiêu phần trăm lợn ở 21 ngày tuổi cho ta khối lượng từ 4,61 kg đến 5,91 kg? (nếu cách tính và vẽ đồ thị minh hoạ)

3. (*5 điểm*) Dựa vào đồ thị và các thông số của đề bài hãy cho biết các giá trị sau đây của mẫu được rút ra từ quần thể nói trên (sử dụng các ký hiệu và các đơn vị đo tương ứng)

a) Trung bình.........………................ b) Độ lệch chuẩn......…...............….............

c) Phương sai...................................... d) Sai số tiêu chuẩn………..........................

e) Hệ số biến động..............………........

2.7. Các thuật ngữ tiếng Anh - Việt

Tiếng Anh	Tiếng Việt	Minitab 12.0	Ký hiệu
Mean	Trung bình	Mean	$\overline{X}$, $\overline{Y}$, μ *
Median	Trung vị	Median	M
Mode	Mode	Mode	Mode
Standard Deviation	Độ lệch chuẩn	StDev	S, σ*
Variance	Phương sai	-	S^2, σ^{2*}
Standard Error	Sai số tiêu chuẩn	SE Mean	SE, $S_{\overline{X}}$, $m_{\overline{X}}$
Variable	Biến	Variable	Var
Maximum	Giá trị lớn nhất	Maximum	Max
Minimum	Giá trị bé nhất	Minimum	Min
Coefficient of Variation	Hệ số biến động	-	Cv

* Các ký hiệu có dấu * trong bảng là các tham số của quần thể

3. Kiểm định giả thiết

3.1. Giả thiết nghiên cứu

3.1.1. Giới thiệu

Ta có thể chia lý thuyết thống kê thành 2 phần lớn:

- Một là, phần thống kê mô tả (như ta đã xem xét ở các phần trước) bao gồm các tóm tắt dưới dạng số, đồ thị … để tóm tắt và mô tả số liệu.
- Hai là, phần suy diễn thống kê, đây là phần rút ra những kết luận về quần thể dựa trên các đại diện mẫu (các số liệu thí nghiệm hay điều tra). Thống kê suy diễn bao gồm:

  Ước tính - các tham số của quần thể như μ, σ từ các đại diện mẫu,

  Kiểm định giả thiết - tiến hành kiểm tra các giả thiết xem các tham số đó xuất phát từ 1 hay từ các quần thể khác nhau.

Ví dụ:

Xem xét đến hiệu lực của một vacxin?

Một phương pháp chăn nuôi mới có làm cho mức độ tăng trọng của lợn nhanh hơn phương pháp hiện tại không?

3.1.2. Giả thiết H_0 và H_1

Trong quá trình nghiên phải tiến hành so sánh sự khác nhau giữa các công thức thí nghiệm (sự tặng trọng của vật nuôi giữa 2 khẩu phần ăn, giữa các giống khác nhau...). Trước khi tiến hành phân tích, đánh giá và đưa ra các kết luận ta phải nêu lên được giả thiết; sau đó tiến hành chứng minh và đưa kết luận, giả thiết đó đúng hay sai ở một mức xác suất nhất định. Một giả thiết như vậy được gọi là *giả thiết H_0*; khi H_0 bị bác bỏ ta phải chọn một giả thiết ngược lại với H_0, đó chính là *đối thuyết H_1*.

3.1.3. Giá trị P

Kiểm định giả thiết dựa trên nguyên tắc xác suất bé; tức là sự kiện không xảy ra sau một lần thí nghiệm. Ta phải chọn một giá trị *P* nhất định để trên cơ sở đó bác bỏ hoặc chấp nhận hoặc bác bỏ H_0. Trong chăn nuôi, thú y ta thường chọn các mức sau 0,05; 0,01; 0,001. *P* chính là xác suất để tồn tại H_0 nếu nó đúng.

3.1.4. Sử dụng giá trị P để rút ra kết luận

Trong thống kê ta thường chọn ngưỡng $P = 0,05$ để làm mức ý nghĩa.

Nếu $P < 0,05 \rightarrow$ giả thiết H_0 bị bác bỏ tức là chấp nhận H_1

Nếu $P \geq 0,05 \rightarrow$ giả thiết H_0 không bị bác bỏ

3.1.5. Sai lầm loại I và loại II

Trong quá trình kiểm định giả thiết ta sẽ chọn H_0 hoặc H_1 tuỳ theo kết quả phân tích số liệu. Như vậy ta có thể mắc phải những sai lầm sau:

- Bác bỏ giả thiết H_0 mặc dù giả thiết đó đúng - **Sai lầm loại I**
- Chấp nhận giả thiết H_0 mặc dù giả thiết đó sai - **Sai lầm loại II**

	Bác bỏ H_0	Chấp nhận H_0
H_0 đúng	Sai lầm loại I	Quyết định đúng
H_0 sai	Quyết định đúng	Sai lầm loại II

3.1.6. Xác suất mắc sai lầm

Chúng ta cần phải hiểu được tầm quan trọng của 2 loại sai lầm này; chúng đóng một vai trò quan trọng trong việc xác định dung lượng mẫu phù hợp nhất đối với một thí nghiệm (chúng ta sẽ xem xét cụ thể hơn ở phần thiết kế thí nghiệm)

- **Xác suất mắc sai lầm loại I** được ký hiệu α. Đây là xác suất mắc sai lầm khi loại bỏ H_0. Giá trị α có thể kiểm tra được vì giá trị này ta tự chọn. Giá trị α được chọn trong quá trình thiết kế thí nghiệm sẽ quyết định việc bác bỏ hay chấp nhận H_0 hay nói một cách khác chúng ta sẽ loại bỏ H_0 nếu $P < \alpha$.
- **Xác suất mắc sai lầm loại II** được ký hiệu β. Đây chính là xác suất không loại bỏ H_0 khi giả thiết này sai. Chúng ta có thể kiểm soát được β bằng cách xem xét các yếu tố làm ảnh hưởng đến β (α, dung lượng mẫu, các yếu tố thí nghiệm, sự biến động của dữ liệu). Trong thực tế ta quan tâm đến hiệu số 1- β; đây chính là độ mạnh của phép thử. 1- β này không bao giờ đạt được 1 (100%); qua các thực nghiệm cho thấy β ít khi vượt quá 0,8 (80%), thí nghiệm có quy mô lớn thì độ mạnh của phép thử càng cao tức là chúng ta có nhiều cơ may hơn để xác định một cách chính xác sự khác nhau giữa các nghiệm thức.

3.2. Kiểm định 1 mẫu

3.2.1. Giới thiệu

Trong chăn nuôi, thú y chúng ta thường xuyên quan tâm đến sự thích nghi của động vật, mức độ tăng trọng của động vật đối với một loại thức ăn mới… tức là ta phải so sánh giá trị trung bình của các thí nghiệm điển hình với các tham số của quần thể (μ, σ^2) để từ đó rút ra được kết luận.

3.2.2. Kiểm định một mẫu bằng phép thử z nếu biết phương sai của quần thể σ^2

Đối với những bài toán so sánh giá trị trung bình của một mẫu khi đã biết được các tham số của quần thể là giá trị trung bình μ và phương sai σ^2; ta sẽ sử dụng phép thử *z*

Ví dụ

Thời gian mang thai của bò có phân bố chuẩn với giá trị trung bình là 285 ngày và độ lệch chuẩn là 10 ngày, dưới dạng rút gọn $y \sim N(285, 10^2)$.

Thời gian mang thai (ngày) của 6 bò của một giống khác được chọn ra là:

307 293 293 283 294 297

Giả sử rằng sự biến động của giống bò mới tương tự so với tiêu chuẩn.

Câu hỏi được đặt ra là: Có sự khác biệt rõ rệt về thời gian mang thai của giống bò mới so với 285 ngày không?

3.2.2.7. Điều kiện cần thiết để thực hiện phép thử:

- Số liệu của mẫu phải có phân bố chuẩn
- Độ lệch chuẩn của mẫu phải đồng nhất so với quần thể

3.2.2.8. Các bước thực hiện

- **Giả thiết:**

H_0 - Giá trị trung bình của quần thể nghiên cứu bằng trung bình của quần thể ban đầu (quần thể rút mẫu)

H_1 - Giá trị trung bình của quần thể nghiên cứu khác so với quần thể ban đầu (quần thể rút mẫu)

- **Kiểm tra sự phân bố của các giá trị quan sát**

Kiểm tra phân bố chuẩn của số liệu bằng cách quan sát biểu đồ tần suất của chúng với sự trợ giúp của phần mềm *Minitab* 12.0.

- **Tính giá trị *z* thực nghiệm**

$$\boxed{z = \frac{(\bar{y} - \mu)}{\sqrt{\sigma^2 / n}} = \frac{(\bar{y} - \mu)}{\mathrm{se}(\bar{y})}}$$

- **Xác định giá trị *P***

Xác định giá trị P bằng cách so sánh giá trị *z* thực nghiệm với phân bố *z*.

- **Rút ra kết luận**

Từ giá trị P thu được từ bảng tính ta có thể rút ra kết luận:

Nếu $P \geq 0,05$ ta không có cơ sở để bác bỏ H_0 tức là chấp nhận H_0

Nếu $P < 0,05$ ta bác bỏ H_0 tức là chấp nhận H_1

Lưu ý: Trong quá trình tính toán bằng tay, ta khó có thể xác định được giá trị *P* chính xác của phép thử. Ta có thể dùng nguyên tắc sau đây để rút ra kết luận *Nếu giá trị Z thực nghiệm lớn hơn giá trị Z lý thuyết ở mức xác suất đã chọn thì giả thiết H_0 bị bác bỏ và ngược lại*

Để minh hoạ cho các bước vừa nêu trên ta tiến hành gải quyết bài toán đã đặt ra

Lời giải

Biết độ lệch chuẩn σ = 10 ngày, sử dụng phép thử *z*

1. Giả thiết

Giả thiết không: $H_0 : \mu = 285$ ngày

Đối thuyết: $H_1 : \mu \neq 285$ ngày

trong đó μ = giá trị trung bình thời gian mang thai của giống mới

$\bar{y} = (307 + 293 + 293 + 283 + 294 + 297) / 6 = 294{,}5$ ngày

2. Kiểm tra sự phân bố chuẩn của số liệu

Kiểm tra phân bố chuẩn của số liệu bằng *Minitab 12*.

3. Tính giá trị z thực nghiệm:

$$\boxed{z = \frac{(\bar{y} - \mu)}{\sqrt{\sigma^2 / n}} = \frac{(\bar{y} - \mu)}{\mathrm{se}(\bar{y})}}$$

Trong ví dụ này ta có

$$z = \frac{294{,}5 - 285}{\sqrt{10^2 / 6}} = 2{,}33$$

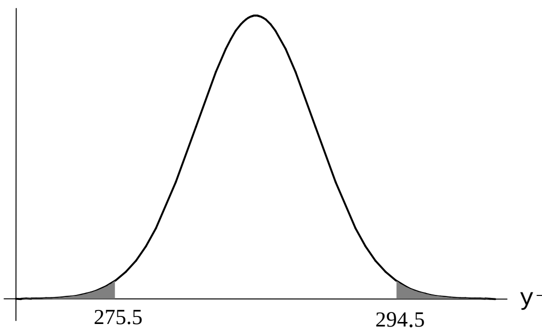
Giả sử rằng giả thiết H_0 đúng (tức là μ = 285 ngày), khi *z* = 2,33 ngày là quan sát từ một phân bố tiêu chuẩn hoá.

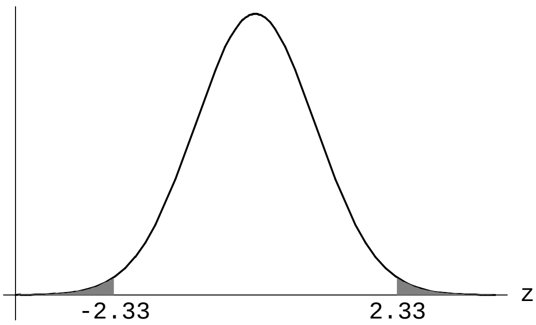
4. Xác định giá trị P

Bây giờ ta sẽ tính xác suất của giá trị z thu được. Giá trị *P* của phép thử là:

$$\begin{aligned} P &= P(\bar{y} < 275{,}5 \text{ hay } \bar{y} > 294{,}5) \\ &= P(Z < -2{,}33 \text{ hay } Z > 2{,}33) \\ &= 2 \times P(Z < -2{,}33) \\ &= 2 \times 0{,}010 = 0{,}020 \end{aligned}$$

Chúng ta cũng có thể dùng bảng ở phần phụ lục để xác định giá trị *P*.





5. Kết luận

Nếu H_0 đúng thì cơ may để thu được giá trị trung bình $\bar{y}$ là 2%. Điều khó có thể xảy ra, vì vậy ta ***bác bỏ giả thiết không***.

<u>Kết luận:</u> Thời gian mang thai của giống bò mới có giá trị trung bình khác biệt có ý nghĩa và lớn hơn 285 ngày**.**

Chú ý:

Theo nguyên tắc chung nếu:

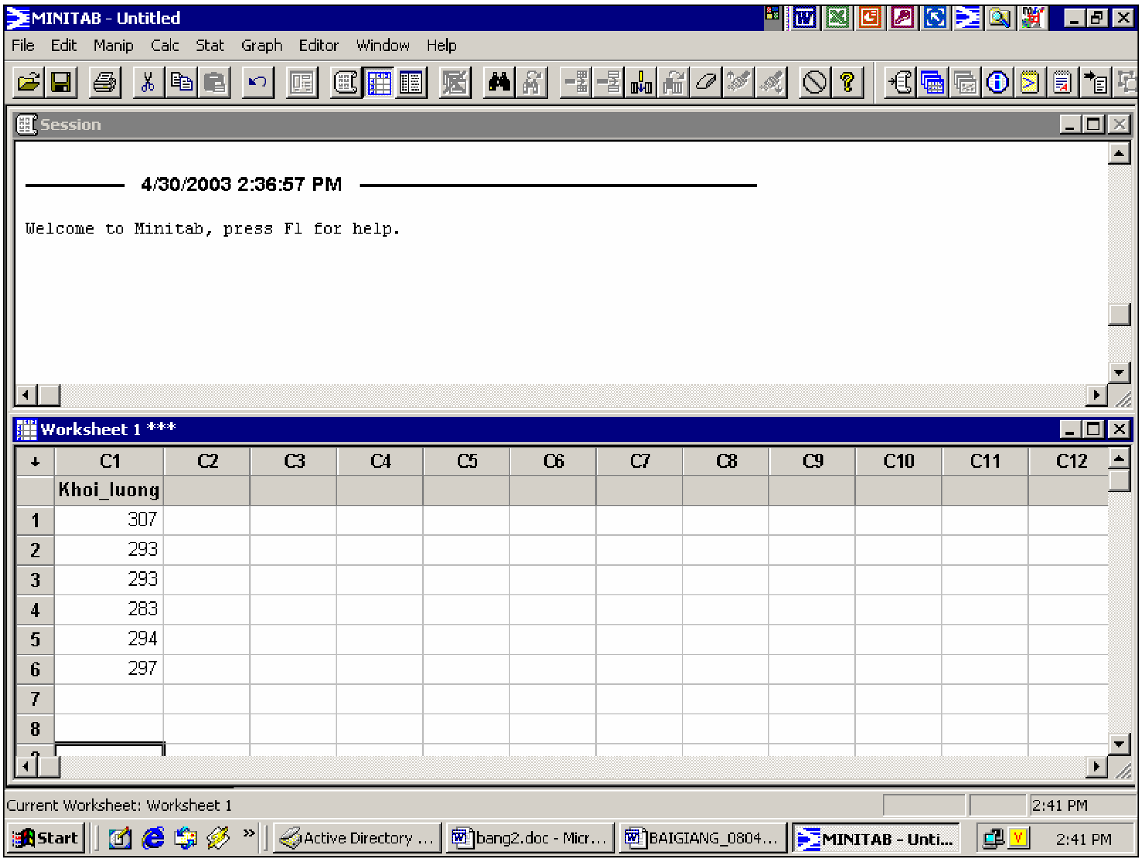
P < 0,05 (bé hơn 1 trên 20) $\Rightarrow$ bác bỏ H_0

P > 0,05 (lớn hơn 1 trên 20) $\Rightarrow$ chấp nhận H_0

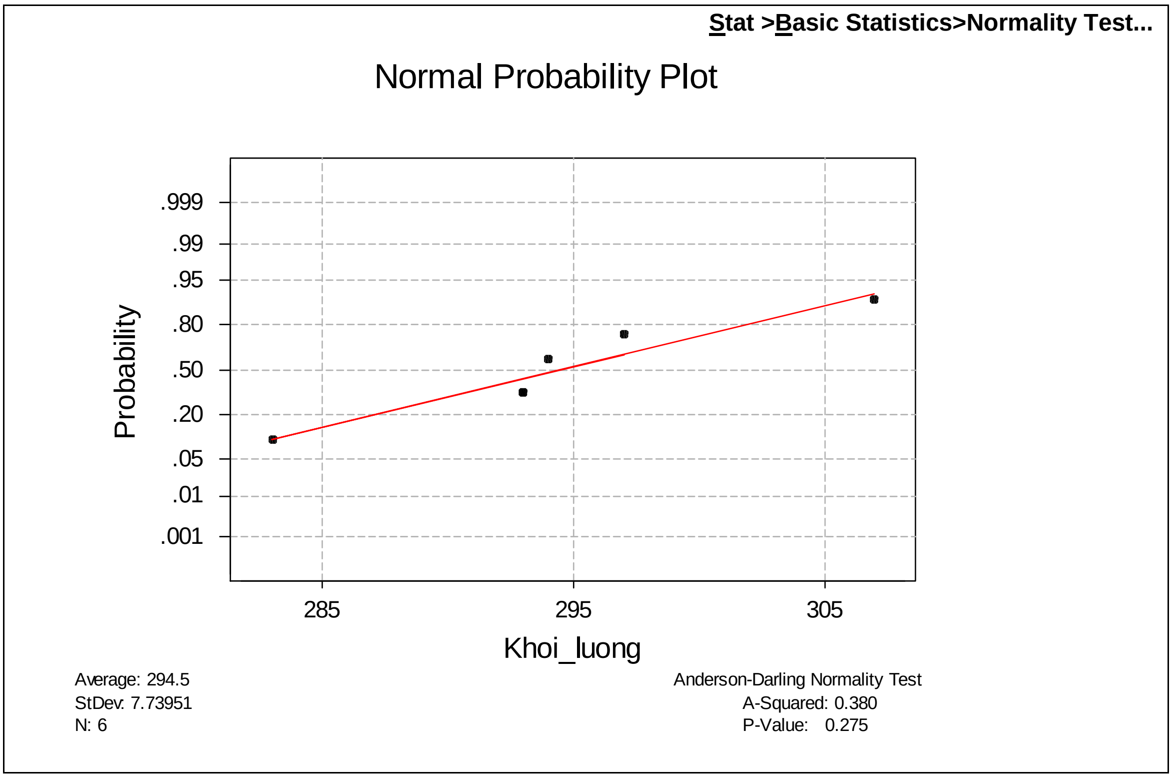
Nếu H_0 được chấp nhận thì không có nghĩa là H_0 hoàn toàn đúng; dung lượng mẫu có thể còn bé để phát hiện ra sự sai khác. Thậm chí ngay cả khi H_0 bị bác bỏ, thì vẫn còn cơ hội rất bé sẽ nằm trong sự sai số. Nếu bạn sử dụng ngưỡng 5%, ***5% kết luận của chúng ta có thể sai khi H_0 đúng!***

Áp dụng phần mềm Minitab

- Nhập số liệu vào Worksheet như hình minh hoạ sau đây, lưu ý rằng dấu phẩy (,) đối với các số thập phân được thay bằng dấu chấm (.); ví dụ 5,3 khi nhập vào *Minitab* là 5.3.



- Kiểm tra phân bố chuẩn của số liệu theo các bước sau đây



Trong kiểm đinh phân bố chuẩn của số liệu thì giả thiết H_0 là số liệu có phân bố chuẩn và đối thuyết H_1 là số liệu không có phân bố chuẩn. Trong ví dụ vừa nêu ta thấy *P*=0,275>0,05, tức là số liệu thoả mãn điều kiện có phân bố chuẩn.

- Tiến hành phân tích số liệu bằng *Minitab*
-

Stat >Basic Statistics>1-Sample Z...

```
MTB > OneZ 'Khoi_luong';
SUBC>   Sigma 10;
SUBC>   Test 285.
```

One-Sample Z: Khoi_luong

```
Test of mu = 285 vs mu not = 285
The assumed sigma = 10

Variable           N      Mean     StDev   SE Mean
Khoi_luong         6    294.50      7.74      4.08

Variable              95.0% CI             Z       P
Khoi_luong    (  286.50,   302.50)      2.33   0.020
```

- Qua phần mềm *Minitab* ta cũng thu được kết quả tương tự như trên. Lưu ý *Minitab* cũng đã tính cho ta khoảng tin cậy 95% là từ 286,5 đến 302,5 ngày; rõ dàng giá trị μ = 285 ngày không nằm trong khoảng tin cậy này.

3.2.3. Kiểm định một mẫu bằng phép thử t

Đối với ví dụ xem ở phần kiểm định *z*, giả sử rằng ta chỉ biết thời gian mang thai trung bình của quần thể μ mà không biết được độ lệch chuẩn của quần thể σ; đối với những trường hợp như vậy ta phải sử dụng phép thử *t* để kiểm định.

Các bước phân tích sẽ thay đổi như thế nào?

Lời giải

Ta **không có** giả thiết σ = 10 ngày, vì vậy sử dụng phép thử *t*

- Giả thiết, $H_0 : \mu = 285$ ngày với đối thuyết $H_1 : \mu \neq 285$ ngày

$\bar{y} = 294{,}5$ ngày và $s = 7{,}74$ ngày.

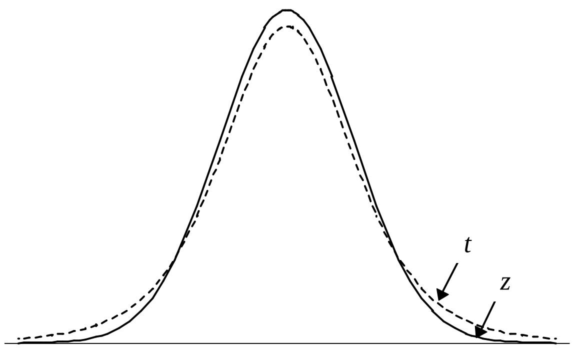
- Tính giá trị *t* thực nghiệm:

$$\boxed{t = \frac{(\bar{y} - \mu)}{\sqrt{s^2 / n}} = \frac{(\bar{y} - \mu)}{\mathrm{se}(\bar{y})}} \text{ với bậc tự do } df = n - 1$$

Như vậy với ví dụ này,

$$t = \frac{294{,}5 - 285}{\sqrt{7{,}74^2 / 6}} = \frac{9{,}5}{3{,}16} = 3{,}01 \text{ với bậc tự do } df = 6 - 1 = 5$$

Giả sử rằng giả H_0 không đúng (tức là μ = 285 ngày), khi $t = 3.01$ là quan sát từ phân bố *t* với bậc tự do $n - 1 = 5$.



Phân bố *t* có các phần đuôi lớn hơn so với phân bố chuẩn. Phân bố này được sử dụng khi độ lệch chuẩn được ước tính từ mẫu. Khi các phần đuôi lớn hơn kéo theo sự sai số lớn hơn trong quá trình ước tính từ phân bố nếu như độ lệch chuẩn của quần thể không biết. Dung lượng mẫu càng lớn thì giá trị độ lệch chuẩn được ước tính càng chính xác hơn cũng như bậc tự do cũng sẽ tăng lên và phân bố *t* dần tiến đến phân bố chuẩn.

Giá trị *P* trong phép thử này là

$$\begin{aligned} P &= P(\bar{y} < 275{,}5 \text{ hay } \bar{y} > 294{,}5) \\ &= P(T_5 < -3{,}01 \text{ hay } T_5 > 3{,}01) \\ &= 2 \times P(T_5 < -3{,}01) \\ &= 2 \times 0{,}015 = 0{,}03 \end{aligned}$$

hoặc từ bảng ta có 0,02 < P < 0,05.

- Kết luận, một lần nữa giá trị P lại nhỏ hơn 0,05, vì vậy chúng ta bác bỏ giả thiết H_0 và kết luận rằng giống bò mới có thời gian mang thai dài hơn.

Chú ý:

Giá trị P trong phép thử t lớn hơn trong phép thử z tức là phép thử t-test không chính xác bằng. Điều có thể giải thích rằng một phần thông tin đã được sử dụng để ước tính giá trị σ của quần thể.

Áp dụng Minitab

```
MTB > OneT 'Khoi_luong';          Stat > Basic Statistics > 1-Sample t...
SUBC>   Test 285.

One-Sample T: Khoi_luong

Test of mu = 285 vs mu not = 285

Variable          N      Mean     StDev   SE Mean
Khoi_luong        6    294.50      7.74      3.16

Variable              95.0% CI            T       P
Khoi_luong    (  286.38,  302.62)      3.01  0.030
```

mChú ý một lần nữa giá trị P, khoảng tin cậy 95% lớn hơn trong phép thử Z nhưng ta vẫn có kết luận tương tự.

3.3. Khoảng tin cậy của trung bình quần thể

3.3.1. Giới thiệu

Kiểm tra giả thiết cho chúng ta biết số liệu có thích hợp với một giá trị trung bình cụ thể μ hay không. Một câu hỏi tiếp theo có thể được đặt ra là:

Miền giá trị nào của giá trị μ
phù hợp với các trung bình quan sát, $\bar{y}$ *?*

Chúng ta cần phải cụ thể hoá mức độ xảy ra hoặc giá trị trung bình của quần thể μ sẽ nằm trong trong khoảng đó.

Để chắc chắn hơn rằng trongkhoảng đó sẽ bao gồm μ*,*
thì giá trị của khoảng đó cũng phải tăng lên.

3.3.2. Công thức tính khoảng tin cậy 95% (95% CI)

Trường hợp 1:	Biết phương sai quần thể σ^2 và cho rằng sự biến động của mẫu là đồng nhất so với tiêu chuẩn, trong trường hợp này chúng ta sử dụng **khoảng tin cậy *z***

$$\boxed{\bar{y} \pm z^{(0{,}025)} \times \sqrt{\sigma^2 / n} = \bar{y} \pm z^{(0{,}025)} \times \mathrm{se}(\bar{y})}$$

trong đó $z^{(0,025)}$ = 1,96 là điểm 2,5% giới hạn trên từ phân bố tiêu chuẩn hoá

Ví dụ

Thời gian mang thai của bò được sử dụng để minh hoạ trong ví dụ. Như ta đã biết thời gian mang thai có phân bố chuẩn là $N(285,10^2)$. Sáu quan sát (n = 6) được rút ra từ một giống bò mới, với thời gian mang thai $\bar{y} = 294{,}5$ ngày.

Lời giải

Nếu biến động của giống mới không hề thay đổi so với tiêu chuẩn, chúng ta chọn σ = 10 ngày; áp dụng công thức tính khoảng tin cậy z

$$\boxed{\bar{y} \pm z^{(0.025)} \times \sqrt{\sigma^2 / n} = \bar{y} \pm z^{(0.025)} \times \mathrm{se}(\bar{y})}$$

Trong ví dụ này,

$$294{,}5 \pm 1{,}96 \times \sqrt{10^2 / 6} = 294{,}5 \pm 8{,}00 = (286{,}5;\ 302{,}5).$$

Như vậy mức độ tin cậy 95% của giá trị trung bình (quần thể) của thời gian mang thai giống bò mới nằm trong trong khoảng từ 286,5 đến 302,5 ngày, mặc dù một giá trị ước tính đơn lẻ tốt nhất là 294,5 ngày.

Trường hợp 2: Không biết phương sai quần thể và cho rằng sự biến động của mẫu quan sát là đồng nhất so với tiêu chuẩn, khi đó ta sẽ ước tính σ^2 từ phương sai của mẫu quan sát s và sử dụng **khoảng tin cậy *t***

$$\boxed{\bar{y} \pm t_{n-1}^{(0.025)} \times \sqrt{s^2 / n} = \bar{y} \pm t_{n-1}^{(0.025)} \times \mathrm{se}(\bar{y})}$$

trong đó $t_{n-1}^{(0.025)}$ là điểm 2,5% của gới hạn trên từ phân bố t với bậc tự do $n - 1$.

Ví dụ

Ta sẽ lấy ví dụ vừa nêu trên để minh hoạ; giả sử ta chỉ biết được thời gian mang thai của bò có phân bố chuẩn với μ = 285 ngày mà không biết phương sai của quần thể. Trong trường hợp này ta sẽ tính khoảng tin cậy t

Lời giải

Phương sai của mẫu là $s^2 = (7{,}74)^2$.

với bậc tự do n – 1 = 6 -1 = 5, điểm 2,5% giới hạn trên của phân bố t là $t_5^{(0.025)}$ = 2,57

Do đó 95% CI là

$$294{,}5 \pm 2{,}57 \times \sqrt{7{,}74^2 / 6} = 294{,}5 \pm 8{,}1 = (286{,}4;\ 302;6).$$

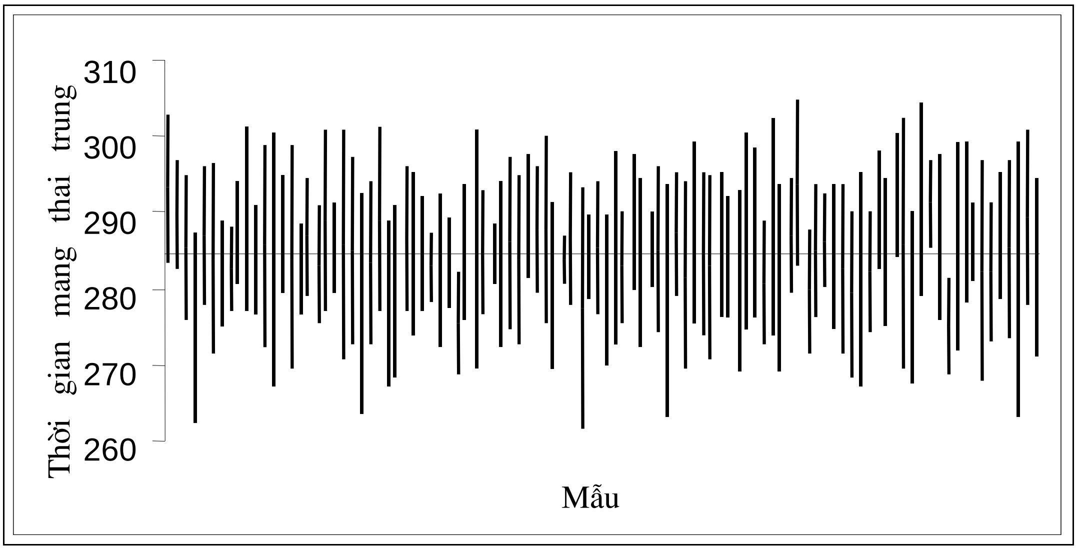
thấy rằng khoảng tin cậy 95% của thời gian mang thai đối với giống mới nằm trong khoảng từ 286,4 đến 302,6 ngày.

Lưu ý:

Động vật thí nKhoảng tin cậy *t* bao giờ cũng lớn khoảng tin cậy *z*; điều này đã được minh chứng rõ trong ví dụ trên.

3.3.3. Ý nghĩa của khoảng tin cậy

Nếu thí nghiệm lặp lại nhiều lần, thì 95% các giá trị trung bình mẫu sẽ rơi vào khoảng tin cậy 95% của quần thể, μ.
gBiểu đồ sau đây sẽ cho ta thấy 100 khoảng tin cậy mô phỏng. Mỗi khoảng tin cậy được xây dựng từ việc rút n = 6 quan sát về thời gian mang thai của bò với giả sử rằng thời gian mang thai có phân bố chuẩn $y \sim N(285, 10^2)$ ngày. Đối với mỗi mẫu, ta tiến hành tính trung bình mẫu $(\bar{y})$ và độ lệch chuẩn (s), sau đó tính khoảng tin cậy 95% theo công thức $(\bar{y} \pm t_{n-1}^{(0.025)} \sqrt{s^2/n})$.



Gần 95% các mẫu mô phỏng này có khoảng tin cây bao gồm giá trị 285. Tuy nhiên trong thực tế chúng ta không biết mẫu nào chứa μ = 285, cũng như ta không biết chính xác μ. Khoảng tin cậy 99% sẽ lớn hơn và chính vì vậy sẽ có nhiều cơ hội có chứa μ

3.4. So sánh 2 mẫu bằng phép thử *t*

3.4.1. Giới thiệu

Trong trường hợp chỉ kiểm định một mẫu (như đã xem xét ở phần 1) , khi so sánh trung bình mẫu $\bar{y}$ với giả thiết trung bình quần thể, μ. Nhưng trong thực tế rất ít có trường hợp như vậy. Thông thường cần có kết luận về mẫu đối với cả 2 quần thể (ví dụ quần thể thứ nhất và thứ hai) và tiến hành so sánh giá trị trung bình của 2 mẫu, giả sử $\bar{y}_1$ và $\bar{y}_2$.

So sánh 2 mẫu bằng phép thử *t* là một trong những phép thử hay được sử dụng trong chăn nuôi và thú y. Phép thử này được sử dụng nhằm so sánh 2 giá trị trung bình từ 2 nhóm độc lập và là mẫu đại diện cho quần thể.

3.4.2. Các điều kiện để tiến hành phép thử

- Động vật thí nghiệm phải được chọn ngẫu nhiên từ quần thể
- Hai mẫu phải độc lập
- Số liệu phải có phân bố chuẩn
- Phương sai giữa 2 mẫu nếu:

  - **Đồng nhất**, chúng ta có thể kiểm tra sự đồng nhất bằng các phép thử phương sai hoặc đơn giản lấy s_1/s_2 (s_1 là độ lệch chuẩn của mẫu 1, s_2 là độ lệh chuẩn của mẫu 2 và giả sử rằng $s_1>s_2$). Nếu tỷ số s_1/s_2 <1,5 thì phương sai có thể coi như là đồng nhất hặc dùng *Minitab*. Nếu các bước vừa nếu trên thoả mãn, ta có thể thực hiện các bước tiếp ở phần 3.3
  - **Không bằng nhau,** thực hiện các bước tiếp theo ở phần 3.4. Tuy nhiên ta cũng có thể tiến hành biến đổi số liệu để đưa các phương sai đồng nhất để sử dụng phép thử ở phần 3.3. Nếu biến đổi số liệu không mang lại những kết quả như mong đợi, ta có thể sử dụng phương pháp thống kê phi tham số để so sánh (sẽ không đề cập trong khoá học này)

3.4.3. Kiểm định 2 mẫu bằng phép thử **t** *(phương sai bằng nhau)*

- Giả thiết

  H_0: Trung bình của 2 quần thể bằng nhau $\mu_1 = \mu_2$

  H_1: Trung bình của 2 quần thể không bằng nhau $\mu_1 \neq \mu_2$

- Kiểm tra phân bố chuẩn của số liệu

Kiểm tra phân bố của số liệu bằng cách quan sát biểu đồ tần suất của chúng với sự trợ giúp của phần mềm *Minitab* 12.0.

- Kiểm tra sự đồng nhất của phương sai
- Tính giá trị *t* thực nghiệm

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{\bar{y}_1 - \bar{y}_2}{\mathrm{se}(\bar{y}_1 - \bar{y}_2)}$$ với bậc tự do $$\begin{aligned} df &= (n_1 - 1) + (n_2 - 1) \\ &= n_1 + n_2 - 2 \end{aligned}$$

trong đó n_1, n_2 là dung lượng mẫu (số quan sát) của mẫu thứ 1 và 2

$\bar{y}_1$ và $\bar{y}_2$ là giá trị trung bình của mẫu thứ 1 và 2

$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$ là phương sai ước tính chung, σ^2

- **Xác định giá trị P**

Xác định giá trị P bằng cách so sánh giá trị *t* thực nghiệm với phân bố *t* vớ bậc tự do là $n_1 + n_2 - 2$ trong bảng t ở phần phụ lục.

- **Rút ra kết luận**

Tuỳ thuộc vào giá trị *P* thu được, ta có thể đưa ra kết luận về giả thiết:

Nếu $P \geq 0,05$ giả thiết H_0 được chấp nhận

Nếu $P < 0,05$ bác bỏ giả thiết H_0 tức là chấp nhận H_1

- **Khoảng tin cậy sự sai khác giữa 2 giá trị trung bình ($\mu_1 - \mu_2$)**

Ước tính tốt nhất cho giá trị trung bình của quần thể μ_1 và μ_2 là các giá trị trung bình mẫu $\bar{y}_1$ và $\bar{y}_2$. Vì vậy ước tính tốt nhất cho sự sai khác $\mu_1 - \mu_2$ chính là $\bar{y}_1 - \bar{y}_2$, được gọi là ***ước lượng điểm***.

Khoảng tin cậy 95% sự sai khác giữa 2 giá trị trung bình được xác định theo công thức sau:

$$\bar{y}_1 - \bar{y}_2 \pm t_{n_1+n_2-2}^{(0.025)} \times \sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \bar{y}_1 - \bar{y}_2 \pm t_{n_1+n_2-2}^{(0.025)} \times \mathrm{se}(\bar{y}_1 - \bar{y}_2)$$

trong đó $t_{n_1+n_2-2}^{(0.025)}$ là 2,5% giá trị phía trên của phân bố *t* với bậc tự do $n_1 + n_2 - 2$.

Ví dụ

Để so sánh khối lượng của 2 giống bò, khối lượng của 12 con bò được chọn ngẫu nhiên đối với giống thứ nhất và 15 con đối với nhóm thứ 2. Khối lượng (kg) của chúng được trình bày ở bảng dưới:

Khối lượng (kg) của 2 giống bò (Campbell, 1989, trang193)

Giống 1	187,6	180,3	198,6	190,7	Giống 2	148,1	146,2	152,8	135,3
	196,3	203,8	190,2	201,0		151,2	146,3	163,5	146,6
	194,7	221,1	186,7	203,1		162,4	140,2	159,4	181,8
						165,1	165,0	141,6	

Câu hỏi dặt ra "*Khối lượng của 2 giống bò có sự sai khác không?*"

Sau đây là các tham số thống kê mô tả từ bộ số liệu trên.

	Giống 1	Giống 2
Trung bình mẫu (kg)	196,2	153,7
Độ lệch chuẩn mẫu (kg)	10,62	12,30

Lời giải

1. Giả thiết $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

2. Kiểm tra phân bố chuẩn của số liệu

Kiểm định phân bố chuẩn của số liệu bằng *Minitab*. Giả sử rằng số liệu có phân bố chuẩn ta sẽ tiến hành bước tiếp theo.

3. Sự đồng nhất của phương sai

Ta có $s_2 / s_1 = 12,30 / 10,62 = 1,16$ ✓ $<1,5$

4. Tính giá trị t thực nghiệm

Ta có $\bar{y}_1 - \bar{y}_2 = 196,2 - 153,7 = 42,5$ kg,

$$s^2 = \frac{11 \times 10,62^2 + 14 \times 12,30^2}{25} = 134,33$$, và $s = \sqrt{134.33} = 11,59$ kg.

Chú ý s là giá trị ước tính giữa 10,62 và 12,30 kg. Ta có thể luôn kiểm tra s chung luôn nằm giữa s_1 và s_2. Sai số tiêu chuẩn của hiệu số giữa các giá trị trung bình là

$$\text{se}(\bar{y}_1 - \bar{y}_2) = \sqrt{134.33\left(\frac{1}{12} + \frac{1}{15}\right)} = 4,489 \text{ kg.}$$

Giá trị t thực nghiệm là

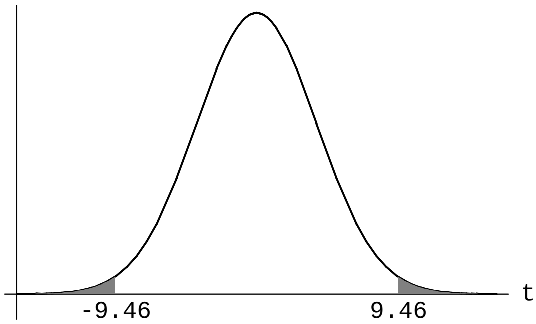
$$t = \frac{\bar{y}_1 - \bar{y}_2}{\text{se}(\bar{y}_1 - \bar{y}_2)} = \frac{42,5}{4,489} = 9,46$$ bậc tự do $df = 12 + 15 - 2 = 25$.

5. Xác định giá trị P

Giả sử rằng giả thiết H_0 đúng ($\mu_1 = \mu_2$), khi $t = 9,46$ là một giá trị quan sát từ phân bố t với bậc tự do là 25. Tra bảng ở phần phụ lục ta thấy $P < 0,001$.

Giá trị P đối với phép thử này là

$$\begin{aligned} P &= P(\bar{y}_1 - \bar{y}_2 < -42,5 \text{ or } \bar{y}_1 - \bar{y}_2 > 42,5) \\ &= P(T_{25} < -9,46 \text{ or } T_{25} > 9,46) \\ &= 2 \times P(T_{25} < -9,46) \\ &= 2 \times 0,0000 = 0,0000. \end{aligned}$$



5. Kết luận

Vì P < 0,001 ta bác bỏ giả thiết H_0 và kết luận rằng trọng lượng của 2 giống bò khác nhau (ở mức P < 0,001). Giống bò thứ nhất nặng hơn giống bò thứ 2 là 42,5 kg.

6. Khoảng tin cậy $\mu_1 - \mu_2$

Ta có, $n_1 + n_2 - 2 = 13 + 15 - 2 = 25$, và $t_{25}^{(0.025)} = 2{,}060$.

Sai số chuẩn là $se(\bar{y}_1 - \bar{y}_2) = 4{,}489$ kg. .

Như vậy khoảng tin cậy 95% $\mu_1 - \mu_2$ là $42{,}5 \pm 2{,}060 \times 4{,}489 = 42{,}5 \pm 9{,}246 = (33{,}2; 51{,}7)$ kg.

Lưu ý rằng khoảng tin cậy này không chứa số 0, với giả thiết không $\mu_1 - \mu_2 = 0$.

Áp dụng Minitab:

Các bước phân tích trên sẽ được thực hiện trong *Minitab*.

Trước hết kiểm tra sự đồng nhất của độ lệch chuẩn

```
MTB > Describe 'P_Giong2' 'P_Giong1'
```

Stat > Basic Statistics > Display Descriptive Statistics...

Descriptive Statistics: P_Giong2, P_Giong1

Variable	N	Mean	Median	TrMean	StDev	SE Mean
P_Giong2	12	196.18	195.50	195.27	10.62	3.06
P_Giong1	15	153.70	151.20	152.95	12.30	3.18

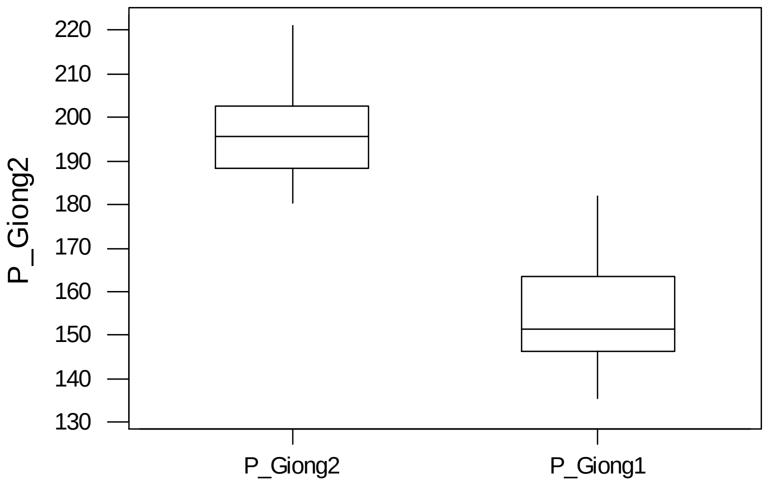
Variable	Minimum	Maximum	Q1	Q3
P_Giong2	180.30	221.10	188.25	202.58
P_Giong1	135.30	181.80	146.20	163.50

Ta thấy tỷ số giữa 2 độ lệch chuẩn là 12,30 / 10,62 < 1,5; như vậy điều kiện 2 phương sai đồng nhất được thoả mãn. Kiểm định *t* phương sai chung có thể sử dụng được (trường hợp tỷ số giữa 2 phương sai lớn hơn 2 ta sẽ xem xét ở phần 1.4.4).

Bây giờ ta sẽ kiểm trả giả thiết về phân bố chuẩn của số liệu. Tốt nhất cho hiển thị số liệu cả hai nhóm đồng thời. Cách này cho ta trực diện có thể kiểm tra được sự đồng nhất của độ lệch chuẩn cũng như phân bố của số liệu.

```
MTB > Boxplot 'P_Giong2' 'P_Giong1';
SUBC>   Box;
SUBC>     Type 0;
SUBC>     Color 0 0;
```

<u>G</u>raph > <u>B</u>oxplot và chọn các options sau
<u>F</u>rame > <u>A</u>xis
<u>F</u>rame > M<u>u</u>ltiple Graphs…> Overlay graphs on the same page
Edit Attributes of ***IQRange Box*** to set ***FillType*** of box as ***None***



Cả hai nhóm cho ta thấy số liệu về trọng lượng có phân bố gần chuẩn, điều cần thiết đối với phép thử *t*. Bây giờ chúng ta tiến hành phép thử đối với giả thiết.

```
MTB > TwoSample 'P_Giong2' 'P_Giong1';
SUBC>   Pooled.
```

<u>S</u>tat > <u>B</u>asic Statistics > <u>2</u>-Sample t... /

Two-Sample T-Test and CI: P_Giong2, P_Giong1

```
Two-sample T for P_Giong2 vs P_Giong1

           N     Mean    StDev   SE Mean
P_Giong2  12    196.2     10.6       3.1
P_Giong1  15    153.7     12.3       3.2

Difference = mu P_Giong2 - mu P_Giong1
Estimate for difference:  42.47
95% CI for difference: (33.23, 51.72)
T-Test of difference = 0 (vs not =): T-Value = 9.46  P-Value = 0.000
DF = 25
Both use Pooled StDev = 11.6
```

Từ kết quả phân tích bằng phần mềm Minitab, ta cũng có các kết luận tương tự

3.4.4. *Kiểm định 2 mẫu bằng phép thử t (phương sai không bằng nhau)*

Khi gặp phải trường hợp 2 mẫu mà phương sai không bằng nhau (tỷ số của độ lệch chuẩn lớn hơn 1,5), ta không thể dùng phương sai ước tính chung cho phép thử như đã trình bày ở phần 3.3. Một số phần mềm thống kê sẽ giúp chúng ta giải quyết những bài toán khi phương sai không đồng nhất, trong đó có phần mềm *Minitab* 12.0.

Nếu bạn muốn thực hiện phép thử bằng cách tính tay, hãy sử dụng các công thức sau đây để thực hiện. Các kỹ thuật tính tay sẽ không được đề cập chi tiết trong khoá học này.

- **Tính *t* thực nghiệm**

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \text{ với bậc tự do } df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Các bước khác được tiến hành tương tự như phần 3.3.

Ví dụ:

Một thí nghiệm sinh lý học động vật nhằm nghiên cứu khả năng hấp thụ của động vật lưỡng cư.   Phần trăm tăng trọng cơ thể sau khi ngâm mình trong nước sau hai giờ được ghi lại sau đối với hai loại động vật (ếch và cóc)

Câu hỏi đặt ra là: Cóc hay ếch hấp thụ nước nhiều hơn?

Số liệu:

Cóc	2,31	25,23	28,37	14,16	28,39	27,94	17,68	
Ếch	0,85	2,90	2,47	17,72	3,82	2,86	13,71	7,38

Chúng tôi chỉ áp dụng phần mềm *Minitab* 12.0 để giải quyết ví dụ này

```
MTB > desc c1;
SUBC> by c2.
```
Stat > Basic Statistics > Display Descriptive Statistics…

Descriptive Statistics: Coc, Ech

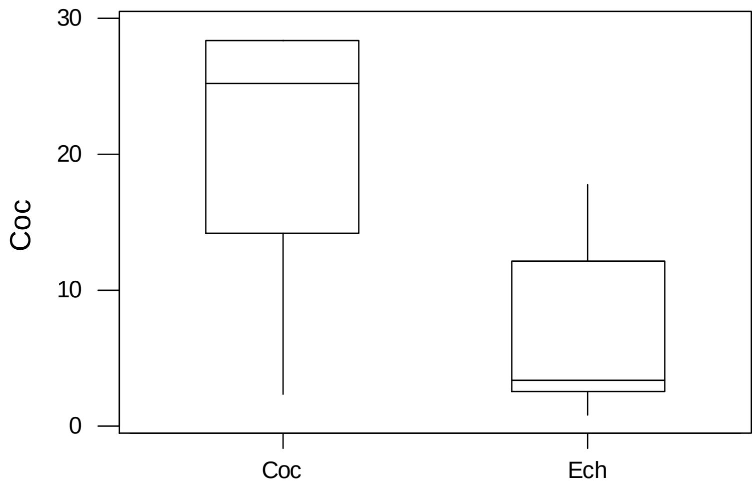
```
Variable             N       Mean     Median     TrMean      StDev    SE Mean
Coc                  7      20.58      25.23      20.58       9.84       3.72
Ech                  8       6.46       3.36       6.46       6.10       2.16

Variable       Minimum    Maximum         Q1         Q3
Coc               2.31      28.39      14.16      28.37
Ech               0.85      17.72       2.57      12.13
```

Boxplot Ech

```
MTB > Boxplot 'Coc' 'Ech';
```

Graph > Boxplot…



Độ lệch chuẩn của hai nhóm rất khác nhau (Ếch: 6,10 so với Cóc: 9,84), như vậy phép thử *t* sử dụng phương sai chung không thể áp dụng được. Biểu đồ hộp cũng chứng tỏ sự biến động cũng không bằng nhau vì vậy ta sẽ sử dụng phép thử *t* của Satterthwaite.

Two-Sample T-Test and CI: Coc, Ech

Stat > Basic Statistics > 2-sample t…

```
Two-sample T for Coc vs Ech

     N      Mean     StDev   SE Mean
Coc  7     20.58      9.84       3.7
Ech  8      6.46      6.10       2.2

Difference = mu Coc - mu Ech
Estimate for difference:  14.12
95% CI for difference: (4.40, 23.84)
T-Test of difference = 0 (vs not =): T-Value = 3.28  P-Value = 0.009  DF = 9
```

Ta thấy rằng, cóc hấp thụ nhiều nước hơn ếch (P = 0,009). Độ tự do trong phép thử này (9) nhỏ hơn trong phép thử *t* chung ($n_1 + n_2 - 2 = 13$)

3.5. So sánh cặp đôi bằng phép thử t

3.5.1. Giới thiệu

So sánh cặp đôi bằng phép thử t được sử khi 2 đại diện mẫu được rút từ quần thể có liên hệ với nhau hoặc những quan sát theo từng cặp. Ví dụ theo dõi 10 con chó, mỗi con nhận được 2 cách xử lý khác nhau.

3.5.2. Các điều kiện để tiến hành phép thử

Các điều kiện cần thiết đối với phép so sánh cặp đôi là:

- Động vật thí nghiệm được chọn một cách ngẫu nhiên
    - Mỗi động vật vừa là đối chứng vừa là thí nghiệm được chọn ngẫu nhiên từ quần thể.
    - Cặp tự nhiên - mỗi cặp được chọn ngẫu nhiên từ quần thể có những đặc điểm tương tự nhau, ví dụ như sinh đôi cùng chứng, cùng giới tính, sinh cùng lứa...
    - Cặp tương đồng - Động vật được chọn ngẫu nhiên từ quần thể, sau đó được phân thành từng cặp tương tự nhau dựa trên các tính trạng khác nhau; ví dụ có cùng độ tuổi, trọng lượng cơ thể, thể trạng, cùng bố...
- Động vật được phân về các lô hoàn toàn ngẫu nhiên
    - Trường hợp động vật thí nghiệm vừa là đối chứng; mỗi động vật được áp dụng một trong hai công thức thí nghiệm hoàn toàn ngẫu nhiên, công thức còn lại sẽ được áp dụng sau đó.
    - Đối với các trường hợp cặp tự nhiên hay tương đồng, mỗi động vật thí nghiệm trong từng cặp được áp dụng công thức thí nghiệm hoàn toàn ngẫu nhiên.
- Sự chênh lệch của từng cặp quan sát phải có phân bố chuẩn hoặc gần chuẩn.

Trường hợp số liệu nêu ở phần trên không có phân chuẩn, chúng ta có thể tiến hành biến đổi số liệu để áp dụng phép thử t cặp đôi. Nếu biến đổi số liệu cũng không mang lại cho ta kết quả mong đợi, thì có thể áp dụng *thống kê phi tham số* để so sánh (sẽ không đề cập đến trong khoá học này)

3.5.3. Các bước giả quyết

- **Giả thiết**

    H_0: Trung bình của sự chênh lệch của từng cặp trong quần thể $\mu_d = 0$

    H_1: Trung bình của sự chênh lệch của từng cặp trong quần thể $\mu_d \neq 0$

- **Kiểm tra phân bố chuẩn của số liệu (phần chênh lệch giữa từng cặp)**

Kiểm tra phân bố của số liệu bằng cách quan sát biểu đồ tần suất của chúng với sự trợ giúp của phần mềm *Minitab* 12.0. Có thể tham khảo thêm ở phần 3.3 và 3.4.

- **Kiểm tra sự đồng nhất của phương sai**

Ưu điểm của phép thử cặp đôi là không phải kiểm tra sự đồng nhất của phương sai; bởi vì ở đây chúng ta chỉ kiểm định chỉ một mẫu đó chính là sự chênh của từng cặp mà thôi.

- Tính giá trị *t* thực nghiệm

$$\boxed{t = \frac{\bar{d}}{SE(\bar{d})} = \frac{\bar{d}}{s_d / \sqrt{n}}} \text{ với } \boxed{df = n - 1}$$

- $SE(\bar{d})$ sai số tiêu chuẩn ước tính của sự chênh lệch
- s_d là sai độ lệch chuẩn ước tính của sự chênh lệch
- n là dung lượng mẫu
- $\bar{d}$ giá trị trung bình chênh lệch của cặp

- **Xác định giá trị *P***

Xác định giá trị *P* bằng cách so sánh giá trị *t* thực nghiệm với phân bố *t* vớ bậc tự do là *n* - 1 trong bảng t ở phần phụ lục hoặc với sự trợ giúp của phần mềm thống kê *Minitab*.

- **Rút ra kết luận**

Tuỳ thuộc vào giá trị *P* thu được, ta có thể đưa ra kết luận về giả thiết:

Nếu $P \geq 0{,}05$ giả thiết H_0 được chấp nhận

Nếu $P < 0{,}05$ bác bỏ giả thiết H_0 tức là chấp nhận H_1

- **Khoảng tin cậy của giá trị trung bình sự chênh lệch**

Khoảng tin cậy 95% của giá trị trung bình được xác định theo công thức sau:

$$\boxed{\bar{d} \pm t_{n-1}^{0{,}025} SE(\bar{d}) = \bar{d} \pm t_{n-1}^{0{,}025} (s_d / \sqrt{n})}$$

trong đó $t_{n_1+n_2-2}^{(0.025)}$ là 2,5% giá trị phía trên của phân bố *t* với bậc tự do *n – 1*.

3.5.4. Ví dụ

Có 15 trại cùng tham gia thử nghiệm khẩu phần ăn của lợn, trong đó khẩu phần ăn bình thường (khẩu phần A) được so sánh với khẩu phần có bổ sung đồng (khẩu phần B). Mỗi trại chọn ra 2 khu nuôi lợn tương tự nhau một cách tối đa về các chỉ tiêu và phân chia theo khẩu phần một cách ngẫu nhiên; một khu cho ăn với một trong hai khẩu phần khu kia cho ăn với khẩu phần còn lại . Tăng trọng trung bình (kg/ngày) của một con lợn đối với mỗi khu được ghi lại như sau:

Trại	Khẩu phần		Trại	Khẩu phần		Trại	Khẩu phần	
	A	B		A	B		A	B
1	0,42	0,53	6	0,50	0,52	11	0,50	0,51
2	0,53	0,47	7	0,44	0,44	12	0,54	0,54
3	0,48	0,56	8	0,45	0,46	13	0,46	0,50
4	0,50	0,59	9	0,30	0,43	14	0,48	0,50
5	0,42	0,47	10	0,52	0,57	15	0,53	0,59

Câu hỏi đặt ra là: *Khi bổ sung thêm đồng có làm cho tăng trọng cao hơn không?*

Lời giải

Từ mỗi trại 2 quan sát được chọn ra, một giá trị trung bình được chọn ra từ mỗi khu. Như vậy thí nghiệm và số liệu thu được được bố trí dưới dạng cặp đôi. Nếu cả thảy có 30 trại được sử dụng để nghiên cứu và chọn ngẫu nhiên 15 trại cho khẩu A và 15 trại còn lại cho khẩu phần B; như vậy các mẫu hoàn toàn độc lập với nhau.

Trong phần này phần tính tay sẽ không đề cập tới. Tuy nhiên các bạn cũng có thể tự tính toán theo các bước đã nêu và hãy so sánh kết quả đạt được với phần mềm *Minitab*.

3.5.5. Áp dụng Minitab

Giả thiết H_0: $\mu_1 = \mu_2$ với H_1: $\mu_1 \neq \mu_2$

Sắp xếp số liệu vào WorkSheets của *Minitab* như sau

Data Display **Manip > Display data...**

```
 Row      A       B

   1   0.42   0.53
   2   0.53   0.47
   3   0.48   0.56
   4   0.50   0.59
   5   0.42   0.47
   6   0.50   0.52
   7   0.44   0.44
   8   0.45   0.46
   9   0.30   0.43
  10   0.52   0.57
  11   0.50   0.51
  12   0.54   0.54
  13   0.46   0.50
  14   0.48   0.50
  15   0.53   0.59
```

Tính các tham số thống kê mô tả

Descriptive Statistics: A, B **Stat > Basic Statistics > Display Discriptive Statistics**

```
Variable            N       Mean     Median     TrMean      StDev    SE Mean
A                  15     0.4713     0.4800     0.4792     0.0614     0.0159
B                  15     0.5120     0.5100     0.5123     0.0517     0.0134

Variable      Minimum    Maximum         Q1         Q3
A              0.3000     0.5400     0.4400     0.5200
B              0.4300     0.5900     0.4700     0.5600
```

Tiến hành phép thử *t* cặp đôi

Paired T-Test and CI: A, B **Stat > Basic Statistics > Paired t…**

```
Paired T for A - B

                 N      Mean     StDev   SE Mean
A               15    0.4713    0.0614    0.0159
B               15    0.5120    0.0517    0.0134
Difference      15   -0.0407    0.0489    0.0126

95% CI for mean difference: (-0.0678, -0.0136)
T-Test of mean difference = 0 (vs not = 0): T-Value = -3.22  P-Value = 0.006
```

Với P = 0,006, ta có thể kết luận rằng khẩu phần có bổ sung đồng đã làm tăng trọng trung bình của lợn tăng lên.

3.6. Bài kiểm tra số 2

A) Trong một thí nghiệm, 100 con cừu được nuôi theo một chế độ riêng. Mục đích của thí nghiệm là xác định xem chế độ nuôi dưỡng này có làm tăng khối lượng 1 năm tuổi của cừu hay không. 100 con cừu này được lấy mẫu từ quần thể có khối lượng trung bình 1 năm tuổi là 30 kg và độ lệch chuẩn là 5 kg. Chế độ nuôi này mang lại giá trị trung bình là 32 kg. (giả sử khối lượng của 100 cừu nói trên có phân phối chuẩn và độ lệch chuẩn đồng nhất với độ lệch chuẩn của quần thể)

1. (1 *điểm*) Anh (chị) hãy tóm tắt các tham số của đề ra bằng các ký hiệu thích hợp cùng với các đơn vị đo tương ứng

2. (4 *điểm*) Nêu giả thiết của phép thử và cho biết kết luận về giả thiết (không cần nêu từng bước tiến hành phép thử).

B) Để so sánh khối lượng sơ sinh giữa 2 giống lợn Landrace và Yorksire nuôi tại trại Mỹ Văn; tiến hành cân khối lượng sơ sinh của 10 lợn Landrace và 18 con của Yorkshire ngay sau khi sinh. Khối lượng sơ sinh trung bình của 10 lợn Landrace là 1,21 kg và độ lệch chuẩn là 0,15 kg; đối với 18 lợn giống Yorkshire có các giá trị tương ứng là 1,30 kg và 0,11 kg.

1. 1. (2 *điểm*) Anh (chị) hãy tóm tắt các tham số của đề ra bằng các ký hiệu thích hợp cùng với các đơn vị đo tương ứng và cho biết có thể dùng phép thử nào để so sánh, vì sao?

2. (3 *điểm*) Nêu giả thiết của phép thử và cho biết kết luận về giả thiết trên ở mức P = 0,05 (không cần nêu từng bước tiến hành phép thử); biết rằng giá trị t của phép thử t = 1,82.

3.7. So sánh nhiều mẫu bằng phân tích phương sai

3.7.1. Giới thiệu

Trong phần 4 chúng ta đã xem xét kiểm định 2 mẫu dựa trên phân bố *t*. Trong các ví dụ ta đã so sánh 2 giá trị trung bình của 2 lô thí nghiệm và các phép thử này chỉ phát huy tác dụng khi thoả mãn hàng loạt các điều kiện. Trong chương này chúng ta sẽ xem xét một phép thử khác dựa trên phân bố *F* để so sánh các phương sai với nhau. Phép thử này được sử dụng để so sánh hai hay nhiều giá trị trung bình với nhau; các tình huống rất hay bắt gặp trong chăn nuôi và thú y.

3.7.2. Phân bố F

3.7.3. Cơ sở lý thuyết

Mở rộng bài toán kiểm định hai mẫu, khi chúng ta cần so sánh sự đồng nhất của nhiều giá trị trung bình thực nghiệm. Ví dụ, chúng ta muốn so sánh ảnh hưởng của 4 khẩu phần ăn khác nhau đối với tăng trọng của gà (và so sánh mức tăng trọng của chúng). Chúng ta có thể sử dụng hàng loạt các phép thử bằng phương pháp thử *t* đối với 2 mẫu để so sánh từng cặp các nghiệm thức. Chúng ta có cả thảy 6 cặp để so sánh:

1 với 2; 1 với 3; 1 với 4;
2 với 3; 2 với 4;
3 với 4;

Vấn đề đặt ra: Mỗi một phép thử có xác suất 5% sai số với kết quả có ý nghĩa. Với sáu lần thử, sẽ có xác suất $1 - (1 - 0{,}05)^6 = 0{,}2654$ sai số từ kết quả có ý nghĩa. Vì vậy chúng ta cần phải có một phương pháp khác để so sánh sự đồng nhất của tất cả các giá trị trung bình của nghiệm thức.

Nếu quan sát các giá trị thu được ta thấy, trong cùng một công thức cũng có sự sai giữa các cá thể (ví dụ sự khác nhau giữa các cá thể trong từng khẩu phần) còn gọi là ***sai số ngẫu nhiên*** và sự sai khác giữa các công thức với nhau gọi là ảnh hưởng của ***nghiệm thức***; ta có thể mô hình hoá như sau:

Tổng toàn bộ biến động = Biến động nghiệm thức + Biến động do sai số ngẫu nhiên

Chúng ta sẽ tiến hành tính các giá trị này như thế nào?

3.7.4. Các điều kiện để tiến hành phép thử

1. Số liệu phải có phân bố chuẩn
   $y_{ij} \sim N(\mu_i,\sigma^2)$, hoặc $\varepsilon_{ij} \sim N(0,\sigma^2)$,
2. Phương sai (quần thể) của các quần thể đồng nhất $\sigma_1 = \sigma_2 = ... = \sigma_t$.
3. Các mẫu độc lập với nhau và được chọn ngẫu nhiên từ một quần thể có phân bố chuẩn

3.7.5. Các bước tiến hành phân tích

- Bước 1

Nêu lên giả thiết nghiên cứu

H_0: Trung bình của các quần thể bằng nhau $\mu_1 = \mu_2 = \ldots = \mu_t$

H_1: Trung bình của các quần thể không bằng nhau

- Bước 2

Kiểm tra phân bố chuẩn của số liệu bằng cách quan sát biểu đồ tần suất của chúng với sự trợ giúp của phần mềm *Minitab* 12.0 hoặc tham khảo ở các mục trên.

- Bước3

Kiểm tra sự đồng nhất của phương sai $(\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2)$.

Sự đồng nhất của phương sai được kiểm tra bằng phép thử Levene; phép thử này có trong hầu hết các phần mềm thống kê không loại trừ *Minitab* 12.0. Phép thử cho phép so sánh 2 hay nhiều phương sai và cho ta biết ngay kết quả.

Tuy nhiên ta cũng có thể dùng phép thử kinh điển như sau để xác định sự đồng nhất của phương sai:

Nếu tỷ số ***độ lệch chuẩn lớn nhất/độ lệch chuẩn nhỏ nhất < 2*** thì cũng chứng tỏ rằng các phương sai đồng nhất

- Bước 4

Sắp xếp số liệu theo từng nghiệm thức và tính tổng cộng theo nghiệm thức (*T*)và tổng số toàn bộ các giá trị quan sát của thí nghiệm (*G*)

- Bước 5

Xây dựng cấu trúc của bảng phân tích phương sai

Nguồn biến động	Bậc tự do (*df*)	Tổng bình phương (*SS*)	Trung bình bình phương (*MS*)	Giá trị *F* quan sát	Giá trị *P* lý thuyết 5%	1%
Nghiệm thức						
Sai số ngẫu nhiên						
Tổng biến động						

- Bước 6

Xác định bậc tự do (*df*) của nghiệm thức, sai số ngẫu nhiên và tổng biến động

- *df* của tổng biến động = n -1
- *df* của nghiệm thức = t -1
- *df* sai số ngẫu nhiên = *df* tổng biến động - *df* nghiệm thức = (n-1) - (t-1) = n - t

- Bước 7

Xác định giá trị hiệu chỉnh (*CF*) và các tổng bình phương (*SS*) từ các giá trị tổng cộng theo nghiệm thức (*T*) và tổng cộng toàn bộ các giá trị quan sát của thí nghiệm (*G*)

- $CF = \frac{G^2}{n}$
- SS toàn bộ quan sát = $\sum_{i=1}^{n} x_i^2 - CF$
- SS nghiệm thức = $\sum_{i=1}^{t} \frac{T_i^2}{r_i} - CF$

➧ Bước 8

Tính các giá trị trung bình bình phương (MS)

- *MS* nghiệm thức = *SS* nghiệm thức / (t-1)
- *MS* sai số ngẫu nhiên = *SS* sai số ngẫu nhiên / (n-t)

➧ Bước 9

Tính giá trị *F* quan sát để kiểm định mức ý nghĩa của nghiệm thức

- *F* = *MS* nghiệm thức / *MS* sai số ngẫu nhiên

➧ Bước 10

Xác định giá trị *F* lý thuyết trong bảng với *df* nghiệm thức = (t - 1) và *df* sai số ngẫu nhiên = (n - t) ở mức ý nghĩa 5% và 1%

➧ Bước 11

Điền toàn bộ các giá trị cần thiết đã tính toàn vào bảng đã thành lập ở bước 2

➧ Bước 12

So sánh giá trị *F* thực nghiệm với giá trị *F* lý thuyết đã nêu ở bước 7 và đưa ra các kết luận về sự sai khác có ý nghĩa giữa các nghiệm thức theo các quy tắc sau đây:

- Nếu giá trị *F* quan sát lớn hơn giá trị *F* lý thuyết ở mức ý nghĩa 1% ta kết luận **có sự sai khác rõ rệt giữa các nghiệm thức.**
- Nếu giá trị *F* quan sát lớn hơn giá trị *F* lý thuyết ở mức ý nghĩa 5% như bé hơn hoặc bằng giá trị *F* lý thuyết ở mức ý nghĩa 1% ta kết luận **có sự sai khác giữa các nghiệm thức.**
- Nếu giá trị *F* quan sát bé hơn hoặc bằng giá trị *F* lý thuyết ở mức ý nghĩa 5% ta kết luận **không có sự sai khác giữa các nghiệm thức.**

Kết quả tính toán các nguồn biến động được trình bày trong bảng ANOVA

Nguồn biến động	Bậc tự do (*df*)	Tổng bình phương (*SS*)	Trung bình bình phương (*MS*)	Giá trị *F* quan sát	Giá trị *P* lý thuyết	
					5%	1%
Nghiệm thức	*t -1*	$SS_{\text{nghiệm thức}}$	$MS_{\text{nghiệm thức}}$	$F_{\text{quan sát}}$	$F_{5\%}$	$F_{1\%}$
Sai số ngẫu nhiên	*n-t*	$SS_{\text{sai số}}$	$MS_{\text{sai số}}$			
Tổng biến động	*n-1*	$SS_{\text{tổng số}}$				

Trung bình bình phương của phần sai số ($MS_{\text{sai số}}$) là giá trị ước tính của σ^2, phương sai của sai số ngẫu nhiên (như ta đã biết $\varepsilon_{ij} \sim N(0,\sigma^2)$).

Trung bình bình phương của nghiệm thức ($MS_{\text{nghiệm thức}}$) cũng là ước tính của σ^2 nếu H_0 đúng, ngược lại nó sẽ lớn hơn σ^2. Nói cách khác nếu H_0 đúng thì giá trị F quan sát (tỷ số phương sai) sẽ có giá trị gần bằng 1.

Giá trị F càng lớn chứng tỏ giả thiết H_0 sai.

- **Khoảng tin cậy**

Minitab không thực hiện được phép thử *t* để so sánh. Tuy nhiên chúng ta hoàn toàn có thể xây dựng khoảng tin cậy 95% đối với sai khác giữa từng cặp giá trị trung bình. Khoảng tin cậy 95% sự sai khác giữa các giá trị trung bình được xác định tương tự như đối với phép so sánh 2 mẫu bằng phép thử *t*. Tuy nhiên phương sai ước tính chung của phép thử này chính là trung bình bình phương của sai số ngẫu nhiên ($MS_{\text{sai số}}$) trong bảng phân tích phương sai và bậc tự do chính bằng bậc tự do của sai số ngẫu nhiên; hoặc cụ thể hơn theo công thức sau:

Sự sai khác bé nhất giữa 2 nhóm bất kỳ (giả sử nhóm thứ nhất và thứ 2) có ý nghĩa thống kê, giá trị *t* được tính như sau:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\mathrm{MS}_{\text{sai so}}\left(\frac{1}{\mathrm{n}_1} + \frac{1}{n_2}\right)}}$$

Trong đó: $\bar{y}_1, \bar{y}_2$ là giá trị trung bình của nhóm 1 v à 2

n_1, n_2 là dung lượng mẫu của nhóm 1 v à 2

Nếu giá trị $t > t^{0.025}_{df\_saiso}$ (được gọi là *t* lý thuyết) thì sự sai khác đó có ý nghĩa tức là trung bình của quần thể 1 khác với trung bình của quần thể 2.

Ta cũng có thể sử dụng sự sai khác bé nhất có ý nghĩa ở mức 5% (*LSD*) được tính như sau để xác định được sự sai khác của các giá trị trung bình

$$LSD(0{,}05) = t^{(0.025)}_{\text{df\_sai so}} \times \sqrt{\mathrm{MS}_{\text{sai so}} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

trong đó $t^{0.025}_{df\_saiso}$ là 2,5% giá trị phía trên của phân bố *t* với bậc tự do của sai số ngẫu nhiên. Nếu khoảng tin cậy này không chứa số 0 thì có sự sai khác có ý nghĩa giữa 2 nhóm ở mức 5%.

3.7.6. Ví dụ

Một thí nghiệm được tiến hành để so sánh mức độ tăng trọng của gà ở 4 khẩu phần ăn khác nhau. 20 con gà đồng đều nhau được phân một cách ngẫu nhiên về một trong 4 khẩu phần ăn. Như vậy ta có 4 nhóm động vật thí nghiệm, mỗi nhóm gồm 5 gà; kết quả thí nghiệm được ghi lại ở bảng sau (đơn vị tăng trọng tính theo g):

Khẩu phần 1	Khẩu phần 2	Khẩu phần 3	Khẩu phần 4
99	61	42	169
88	112	97	137
76	30	81	169
38	89	95	85
94	63	92	154

Lời giải:

Các tham số thống kê mô tả được trình bày ở bảng sau:

Khẩu phần	1	2	3	4	Tổng
Trung bình mẫu	$\bar{y}_1 = 79{,}0$	$\bar{y}_2 = 71{,}0$	$\bar{y}_3 = 81{,}4$	$\bar{y}_4 = 142{,}8$	$\bar{y}_{\bullet} = 93{,}55$
Độ lệch chuẩn mẫu	$s_1 = 24{,}5$	$s_2 = 31{,}0$	$s_3 = 22{,}9$	$s_4 = 34{,}9$	$s_y = 39{,}52$
Dung lượng mẫu	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$n_4 = 5$	$n = 20$

Tiến hành từng bước như đã nêu ở mục 5.3 để rút ra kết luận

- Bước 1

  H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

  H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

- Bước 2

Kiểm tra phân bố chuẩn của số liệu bằng cách quan sát biểu đồ tần suất của chúng với sự trợ giúp của phần mềm *Minitab* 12.0.

- Bước3

Kiểm tra sự đồng nhất của phương sai $(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2)$.

$\frac{s_4}{s_3} = \frac{34{,}9}{22{,}9} = 1{,}52 < 2$ chứng tỏ rằng các phương sai đồng nhất hoặc dùng phần mềm *Minitab* để kiểm tra

▶ Bước 4

Sắp xếp số liệu theo từng nghiệm thức và tính tổng cộng theo nghiệm thức (*T*)và tổng số toàn bộ các giá trị quan sát của thí nghiệm (*G*)

	Khẩu phần 1	Khẩu phần 2	Khẩu phần 3	Khẩu phần 4
	99	61	42	169
	88	112	97	137
	76	30	81	169
	38	89	95	85
	94	63	92	154
T	**395**	**355**	**407**	**714**
G	395 + 355 + 407 + 714 =**1881**			

▶ Bước 5

Xây dựng cấu trúc của bảng phân tích phương sai

Nguồn biến động	Bậc tự do (*df*)	Tổng bình phương (*SS*)	Trung bình bình phương (*MS*)	Giá trị *F* quan sát	Giá trị *P* lý thuyết 5%	1%
Khẩu phần						
Sai số ngẫu nhiên						
Tổng biến động						

▶ Bước 6

Xác định bậc tự do (*df*) của nghiệm thức, sai số ngẫu nhiên và tổng biến động

- *df* của tổng biến động = *n* -1 = 20 -1 = 19
- *df* của nghiệm thức = *t* -1 = 4 -1 = 3
- *df* sai số ngẫu nhiên = *df* tổng biến động - *df* nghiệm thức = (*n*-1) - (*t*-1) = n - t = 19 -3 = 20 -4 = 16

▶ Bước 7

Xác định giá trị hiệu chỉnh (*CF*) và các tổng bình phương (*SS*) từ các giá trị tổng cộng theo nghiệm thức (*T*) và tổng cộng toàn bộ các giá trị quan sát của thí nghiệm (*G*)

- $CF = \dfrac{G^2}{n} = \dfrac{1881^2}{20} = 176908{,}05$
- *SS* toàn bộ quan sát = $\sum_{i=1}^{n} x_i^2 - CF =$
- = ($99^2 + 88^2 + 76^2 + 38^2 + 94^2 + 61^2 + ...+ 154^2$) - 176908,05 = 29.679
- *SS* nghiệm thức = $\sum_{i=1}^{t} \dfrac{T_i^2}{r_i} - CF =$

$$= \left(\frac{396^2}{5} + \frac{357^2}{5} + \frac{410^2}{5} + \frac{718^2}{5}\right) - 176908{,}05 = 16.467$$

- *SS* sai số = *SS* toàn bộ quan sát - *SS* nghiệm thức = 29.679 – 14.467 = 13.212

- Bước 8

Tính các giá trị trung bình bình phương (MS)

- MS nghiệm thức = SS nghiệm thức / (t-1) = $\frac{16.467}{4-1} = 5.489$
- MS sai số ngẫu nhiên = SS sai số ngẫu nhiên / (n-t) = $\frac{13.212}{20-4} = 826$

- Bước 9

Tính giá trị F quan sát để kiểm định mức ý nghĩa của nghiệm thức

- F = MS nghiệm thức / MS sai số ngẫu nhiên = $\frac{5.489}{826}$ = 6,65

- Bước 10

Tra bảng ở phần phụ lục ta có F lý thuyết trong bảng với df nghiệm thức = 3 và df sai số ngẫu nhiên = 16 ta có: ở mức ý nghĩa 5% và 1% thì các giá trị tương ứng là F = 3,10 và F = 4,94

- Bước 11

Nguồn biến động	Bậc tự do (*df*)	Tổng bình phương (*SS*)	Trung bình bình phương (*MS*)	Giá trị *F* quan sát	Giá trị *P* lý thuyết	
					5%	1%
Khẩu phần	3	16,467	5,489	6,65	3,10	4,94
Sai số ngẫu nhiên	16	13,212	826			
Tổng biến động	19	29,679				

- Bước 12

Ta có thể kết luận rằng, có sự sai khác giữa tăng trọng của các khẩu phần ăn khác nhau (P<0,05) và ở khẩu phần 4 cho ta giá trị tăng trọng lớn nhất.

Áp dụng Minitab

Hiển thị số liệu và các tham số thống kê mô tả

```
MTB > Print 'KP1'-'KP4'.              Manip > Display Data...
```

Data Display

```
 Row    KP1    KP2    KP3    KP4

   1     99     61     42    169
   2     88    112     97    137
   3     76     30     81    169
   4     38     89     95     85
   5     94     63     92    154

MTB > Describe 'KP1'-'KP4'.           Stat > Basic Statistics > Display Descriptive Statistics...
```

Descriptive Statistics: KP1, KP2, KP3, KP4

```
Variable           N      Mean    Median    TrMean     StDev   SE Mean
KP1                5      79.0      88.0      79.0      24.5      10.9
KP2                5      71.0      63.0      71.0      31.0      13.9
KP3                5      81.4      92.0      81.4      22.9      10.2
KP4                5     142.8     154.0     142.8      34.9      15.6

Variable     Minimum   Maximum        Q1        Q3
KP1             38.0      99.0      57.0      96.5
KP2             30.0     112.0      45.5     100.5
KP3             42.0      97.0      61.5      96.0
KP4             85.0     169.0     111.0     169.0
```

Để tiến hành các bước phân tích phương sai ta phải sắp xếp lai số liệu thành 2 cột; một cột thể hiện mức tăng trọng, một cột thể hiện khẩu phần ăn tương ứng theo các bước sau đây:

```
MTB > Name c5 = 'P' c6 = 'KP'      Manip > Stack/Unstack > Stack Columns…
MTB > Stack 'KP1'-'KP4' 'P';
SUBC>   Subscripts 'KP';
SUBC>   UseNames.
MTB >
```

Và tiến hành phân tích phương sai

One-way ANOVA: P versus KP **Stat > ANOVA > One-way (Unstacked)…**

```
Analysis of Variance for P
Source     DF        SS        MS        F        P
KP          3     16467      5489     6.65    0.004
Error      16     13212       826
Total      19     29679
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  --------+---------+---------+--------
KP1         5     79.00     24.47     (-------*------)
KP2         5     71.00     31.02   (------*-------)
KP3         5     81.40     22.88     (-------*-------)
KP4         5    142.80     34.90                      (-------*-------)
                                   --------+---------+---------+--------
Pooled StDev =    28.74                   70       105       140
MTB >
```

Với *P* = 0,004, giả thiết H_0 bị bác bỏ hay nói cách khác hoàn toàn có thể loại bỏ giả thiết rằng tăng trọng trung bình giữa các khẩu phần ăn là bằng nhau.

3.7.7. So sánh giữa các nghiệm thức theo từng cặp

So sánh tăng trọng của chuột ở 4 khẩu phần ăn khác nhau (khẩu phần 1, 2, 3 và 4). Số chuột tham gia vào thí nghiệm vào từng khẩu phần là 7, 8, 6 và 8. Số liệu thu được trình bày ở bảng sau (% tăng trọng so với khối lượng cơ thể):

1	2	3	4
3,42	3,17	3,34	3,64
3,96	3,63	3,72	3,93
3,87	3,38	3,81	3,77
4,19	3,47	3,66	4,18
3,58	3,39	3,55	4,21
3,76	3,41	3,51	3,88
3,84	3,55		3,96
	3,44		3,91

Bài giải: (Dùng phần mềm *Minitab* để giải quyết).

Nhập số liệu vào *Minitab*, tính các tham số thống kê mô tả ta thu được kết quả sau:

```
Variable           N       Mean     Median     TrMean      StDev    SE Mean
1                  7     3.8029     3.8400     3.8029     0.2512     0.0949
2                  8     3.4300     3.4250     3.4300     0.1353     0.0478
3                  6     3.5983     3.6050     3.5983     0.1675     0.0684
4                  8     3.9350     3.9200     3.9350     0.1906     0.0674

Variable     Minimum    Maximum         Q1         Q3
1             3.4200     4.1900     3.5800     3.9600
2             3.1700     3.6300     3.3825     3.5300
3             3.3400     3.8100     3.4675     3.7425
4             3.6400     4.2100     3.7975     4.1250
```

Giả thiết H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (bằng lời, bạn đọc tự nêu)

H_1: $\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$

So sánh sự đồng nhất của phương sai: 0,2512 / 0,1353 = 1,86 < 2

Kiểm tra phân bố chuẩn: bằng cách kiểm tra phân bố chuẩn của sai số ngẫu nhiên (phần dư). Đây là một thí nghiệm mà số động vật tham gia vào từng công thức thí nghiệm hạn chế ($n_1 = 7$, $n_2 = 8$ $n_3 = 6$ và $n_4 = 8$), vì vậy ta không kiểm tra phân bố chuẩn của từng biến riêng biệt. dùng *Minitab* để kiểm tra phân bố chuẩn ta có P = 0,55

Phân tích phương sai

```
Analysis of Variance for P
Source     DF        SS        MS        F        P
KP          3    1.1601    0.3867    10.73    0.000
Error      25    0.9012    0.0360
Total      28    2.0613
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+-------
1           7    3.8029    0.2512                 (-----*-----)
2           8    3.4300    0.1353   (----*-----)
3           6    3.5983    0.1675         (-----*-----)
4           8    3.9350    0.1906                       (----*-----)
                                   ---------+---------+---------+-------
Pooled StDev =   0.1899                   3.50      3.75      4.00
```

Kết luận

Vì P = 0,000 < 0,05 ta bác bỏ H_0 và chấp nhận H_1 (bằng lời, bạn đọc tự nêu)

Trong trường hợp bác bỏ H_0 tức là có ít nhất một giá trị trung bình sai khác có ý nghĩa thống kê. Để biết cụ thể ta tiến hành so sánh từng cặp giá trị trung bình với nhau.

Dùng menu *Comparisons* của *Minitab* ta có

```
Tukey's pairwise comparisons
    Family error rate = 0.0500
Individual error rate = 0.0109
Critical value = 3.89
Intervals for (column level mean) - (row level mean)

                 1           2           3

       2      0.1026
              0.6431

       3     -0.0860     -0.4504
              0.4951      0.1137

       4     -0.4024     -0.7661     -0.6187
              0.1381     -0.2439     -0.0546
```

Nếu nhìn vào Ma trận trên ta thấy $\mu_1 \neq \mu_2$, $\mu_1 = \mu_3$, $\mu_1 = \mu_4$, $\mu_2 = \mu_3$, $\mu_2 \neq \mu_4$, $\mu_3 \neq \mu_4$.

Ta có thể xây dựng một bảng có các chữ cái a, b, c... để thể hiện sự sai khác giữa các nghiệm thức.

Thực hiện theo các bước sau:

- Sắp xếp các giá trị trung bình theo thứ tự giảm dần như sau:

Khẩu phần	Trung bình
1	3,8029
2	3,4300
3	3,5983
4	3,9350

→

Khẩu phần	Trung bình
4	3,9350
1	3,8029
3	3,5983
2	3,4300

- Dựa vào ma trận ở trang 50 để tạo các đường gạch chung cho các khẩu phần có giá trị trung bình bằng nhau; cụ thể như sau:

Khẩu phần	Trung bình			
		a		
4	3,9350		b	
1	3,8029			c
3	3,5983			
2	3,4300			

mỗi một đường thẳng tương ứng với một chữ cái (a, b, c...)

- Từ bảng trên, ta có thể đặt các chữ cái bên cạnh các số trung bình như sau:

Khẩu phần	Trung bình
4	3,9350^{a}
1	3,8029^{ab}
3	3,5983^{bc}
2	3,4300^{c}

- Sắp xếp khẩu phần theo thứ tự tăng dần như ban đầu ta có:

Khẩu phần	Trung bình
1	3,8029^{ab}
2	3,4300^{c}
3	3,5983^{bc}
4	3,9350^{a}

3.8. Bài kiểm tra số 3

Để so sánh khối lượng trứng của 4 giống gà (Hyline, Lương Phượng, Sacsso và 707) nuôi tại trại Quang Trung, ĐH Nông nghiệp I Hà Nội; tiến hành rút ngẫu nhiên và cân khối lượng của 15 quả trứng đối với từng giống. Số liệu thu được trình bày ở bảng bên (đơn vị tính - g). LP - Lương phượng, HL - Hyline, SS - Sacsso, 707 - 707. Anh (chị) có kết luận gì về khối lượng trứng của 4 giống gà nêu trên.

STT	LP	HL	SS	707
1	49,45	51,62	50,45	58,34
2	51,96	57,73	53,51	55,74
3	51,72	53,44	50,12	59,25
4	57,47	54,99	53,91	55,74
5	53,59	48,08	53,95	55,35
6	57,06	56,48	54,70	58,35
7	56,51	51,43	55,43	58,98
8	53,07	54,49	57,20	56,30
9	50,28	56,98	49,21	61,64
10	49,62	50,42	51,10	51,14
11	58,43	53,82	46,94	53,02
12	49,79	48,39	56,74	53,21
13	58,58	47,16	52,51	55,81
14	55,76	49,79	53,24	57,63
15	48,44	51,30	51,54	58,13

3.9. Kiểm định khi bình phương và so sánh các tỷ lệ

3.9.1. Các vấn đề sẽ đề cập tới

- Kiểm định 1 tỷ lệ
- So sánh 2 tỷ lệ
- Bảng tương liên 2×2
- Sử dụng kiểm định khi bình phương để phân tích số liệu trong bảng tương liên 2×2

3.9.2. Giới thiệu

Ở các phần trước ta đã tiến hành kiểm định các giá trị trung bình của các biến liên tục. Trong phần này chúng ta sẽ tiến hành nghiên cứu và kiểm định các tỷ lệ, tức là các biến có phân bố nhị thức và các biến định tính ở nhiều mức độ khác nhau.

Các biến có phân bố nhị thức là những biến định tính ở 2 mức, thông thường được gọi là sự kiện xảy ra và không xảy ra. Ví dụ, Little và cộng sự (1980) đã tiến hành điều tra ảnh hưởng của nhiễm trùng Leptospira đến tỷ lệ sẩy thai ở bò.

3.10. Kiểm định một tỷ lệ

3.10.1. Cơ sở lý thuyết

Như chúng ta đã biết, phân bố mẫu của một tỷ lệ sẽ tiến gần đến phân bố chuẩn khi dung lượng mẫu n lớn; như vậy tỷ lệ ước tính p của mẫu cũng sẽ tiến gần đến tỷ lệ π của quần thể và sai số tiêu chuẩn của mẫu được ước tính là $\sqrt{p(1-p)/n}$. Chúng ta sẽ sử dụng những tính chất này để tiến hành kiểm định tỷ lệ xảy ra của một mẫu trong một quần thể theo các bước sau đây:

Giả thiết

H_0: Tỷ lệ của sự kiện xảy ra trong quần thể bằng một đại lượng π

H_1: Tỷ lệ của sự kiện xảy ra trong quần thể không bằng một đại lượng π

Thu thập số liệu

Mẫu được chọn ngẫu nhiên từ quần thể, sau đó phân loại từng cá thể theo sự kiện xảy hoặc không xảy ra.

Tính giá trị *z* thực nghiệm

$$z = \frac{|p-\pi| - \frac{1}{2n}}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

- p Tỷ lệ quan sát của sự kiện xảy ra
- n Dung lượng mẫu
- $1/2n$ Hệ số hiệu chỉnh
- π Giá trị giả thiết

Xác định giá trị P

Xác định giá trị *P* bằng cách sử dụng bảng phân bố tiêu chuẩn hoá (xem bảng ở phần phụ lục hoặc sử dụng *Minitab*).

Rút ra kết luận

Tuỳ thuộc vào giá trị *P* thu được, ta có thể đưa ra kết luận về giả thiết:

Nếu $P \geq 0{,}05$ giả thiết H_0 được chấp nhận

Nếu $P < 0{,}05$ bác bỏ giả thiết H_0 tức là chấp nhận H_1

Khoảng tin cậy 95% đối tỷ lệ của sự kiện xảy

Khoảng tin cậy 95% đối tỷ lệ π được tính theo công thức sau đây:

$$p \pm 1{,}96\sqrt{\frac{p(1-p)}{n}}$$

3.10.2. Ví dụ

Giả sử chúng ta điều tra giới tính của một quần thể nào đó. Trong một mùa nhất định trong năm người ta thấy tỷ lệ giới tính lúc sinh ra có xu hướng con cái cao hơn. Để giải đáp câu hỏi trên người ta đã tiến hành chọn ngẫu nhiên 297 con chim mới sinh thì thấy có 167 con cái. Liệu có yếu tố nào làm ảnh hưởng đến tỷ lệ giới tính hay không?

Ta áp dụng các bước phân tích như đã nêu ở mục 6.3.1 để giải bài toán này

Giả thiết

H_0: Tỷ lệ giữa số con cái và con đực mới sinh trong quần thể là 0,5

H_1: Tỷ lệ giữa số con cái và con đực mới sinh trong quần thể khác 0,5

Tính tỷ lệ

Tỷ lệ cái trong số 297 con mới sinh ra là 167/297 = 0,562.

Tính giá trị *t* thực nghiệm

$$z = \frac{|p-\pi| - \frac{1}{2n}}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

- *p* Tỷ lệ quan sát của sự kiện xảy ra
- *n* Dung lượng mẫu
- *1/2n* Hệ số hiệu chỉnh
- *π* Giá trị giả thiết

$$z = \frac{|p-\pi| - \frac{1}{2n}}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{|0{,}562-0{,}5| - \frac{1}{2\times 297}}{\sqrt{\frac{0{,}5[1-0{,}5]}{297}}} = 2{,}08$$

Xác định giá trị *P*

Sử dụng bảng phân bố tiêu chuẩn hoá hoặc tính trong *Minitab* ta có *P* = 0,0375.

Rút ra kết luận

$P = 0{,}0375 < 0{,}05$ ta bác bỏ giả thiết H_0

Khoảng tin cậy 95% đối tỷ lệ số con cái mới sinh ra

$$p \pm 1{,}96\sqrt{\frac{p(1-p)}{n}} = 0{,}562 \pm 1{,}96\sqrt{\frac{0{,}562(1-0{,}562)}{297}}$$ tức là từ 0,51 đến 0,62

3.11. So sánh 2 tỷ lệ (các mẫu độc lập)

3.11.1. Giới thiệu

Khi ta tiến hành rút 2 mẫu từ 2 quần thể và sử dụng các tỷ lệ p_1 và p_2 của mẫu để ước tính các tỷ lệ quần thể π_1 và π_2. Ta có thể kiểm định sự đồng nhất của các tỷ lệ quần thể theo 2 cách sau:

- Sử dụng phép thử khi bình phương (χ^2)
- Sử dụng phân bố gần chuẩn đối với phân bố nhị thức

3.11.2. Bảng tương liên 2×2

Hawkins và cộng sự (1993) đã tiến hành nghiêm cứu ảnh hưởng của việc thiến đến sự xuất hiện bệnh tiểu đường ở chuột. Chuột thí nghiệm được chia một cách hoàn toàn ngẫu nhiên về 1 trong 2 cách sử lý thiến và không thiến. Tác giả đã tiến hành so sánh tỷ lệ chuột mắc bệnh tiểu đường trong hai nhóm động vật thí nghiệm độc lập.

Nếu biểu diễn tần suất của 2 sự kiện và 2 cách sử lý trong một bảng; được gọi là bảng tương liên, mỗi hàng thể hiện tần xuất của một sự kiện (chuột mắc bệnh hoặc không) và mỗi cột thể hiện một trong 2 nhóm động vật thí nghiệm (thiến hoặc không). Một bảng tương liên như vậy thường được gọi là bảng tương liên 2×2, bởi vì có 2 hàng và 2 cột. Bảng tương liên r×c là bảng có r hàng và c cột. Ta có thể biểu diễn tần suất quan sát dưới dạng tổng quát sau:

	Nhóm		Tổng số theo hàng
	1	2	
Sự kiện xảy ra	a	b	a + b
Sự kiện không xảy ra	c	d	c + d
Tổng số theo cột	$a + c$	$b + d$	Tổng số $n = a + b + c + d$
Tỷ lệ quan sát của sự kiện xảy ra	$p_1 = \frac{a}{a+c}$	$p_2 = \frac{b}{b+d}$	$p = \frac{a+b}{a+b+c+d}$

3.11.3. Áp dụng χ^2 để so sánh 2 tỷ lệ trong bảng tương liên 2×2

- Cơ sở lý luận

Giả sử không có mối liên hệ nào giữa nhóm và sự kiện, thì ta có thể ước tính được tỷ lệ sự kiện xảy ra giữa 2 nhóm là như nhau. Giả sử ta muốn so sánh 2 tỷ lệ bằng cách tiến hành nghiên cứu mối liên hệ giữa 2 yếu tố như đã nêu ở phần 6.4.2. Yếu tố ở đây chính là một biến với nhiều các cấp hạng phân loại khác nhau. Giả thiết H_0 của chúng ta nêu ra là không có mối liên hệ nào giữa 2 yếu tố; hay nói một cách khác là tỷ lệ của 2 quần thể bằng nhau.

Để kiểm định giả thiết này chúng ta sẽ tiến hành so sánh tần suất quan sát trong mỗi ô của bảng tương liên với tần suất ước tính nếu giả thiết H_0 đúng. Giả thiết H_0 là tỷ lệ sự kiện xảy ra của 2 quần thể bằng nhau. Nếu giả thiết H_0 đúng thì ta có thể ước tính được tỷ lệ chung cho cả 2 quần thể là (a + b)/n để áp dụng cho cả 2 nhóm. Tỷ lệ ước tính cho từng nhóm cũng có thể được tính lần lượt là (a + c)×(a + b)/n và (b + d)×(a + b)/n tương ứng với nhóm 1 và 2. Các giá trị ước tính cho từng ô được thể hiện ở bảng sau:

	Nhóm		Tổng số theo hàng
	1	2	
Sự kiện xảy ra	$\frac{(a+c)(a+b)}{n}$	$\frac{(b+d)(a+b)}{n}$	a + b
Sự kiện không xảy ra	$\frac{(a+c)(c+d)}{n}$	$\frac{(b+c)(c+d)}{n}$	c + d
Tổng số theo cột	$a + c$	$b + d$	Tổng số $n = a + b + c + d$
Tỷ lệ quan sát của sự kiện xảy ra	$p_1 = \frac{a}{a+c}$	$p_2 = \frac{b}{b+d}$	$p = \frac{a+b}{a+b+c+d}$

- Giả thiết
  - Các cá thể được rút một cách ngẫu nhiên từ quần thể
  - Các cá thể thí nghiệm được chia về các cách xử lý hoàn toàn ngẫu nhiên
  - Số liệu được thu thập dưới dạng tần suất (sự kiện xảy ra hoặc không) đối với từng nhóm
  - Tần xuất ước tính trong một ô bất kỳ không được bé hơn 5

3.11.4. Các bước tiến hành

1. Giả thiết H_0: Tỷ lệ sự kiện xảy ra của 2 quần thể bằng nhau,
H_1: Không có mối liên hệ nào giữa 2 yếu tố nghiên cứu

2. Thu thập số liệu và nhập tần suất quan sát vào bảng tương liên 2×2

3. Tính giá trị χ^2 thực nghiệm theo công thức sau

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

4. Xác định bậc tự do (df) của phép thử χ^2 dưới dạng tổng quát, df = (Số hàng - 1) × (Số cột – 1)

5. Xác định giá trị *P* bằng cách so sánh giá trị χ^2 thực nghiệm với phân bố χ^2 trong phần phụ lục với bậc tự do df = (Số hàng - 1) × (Số cột – 1)

6. Rút ra kết luận

Nếu $P \geq 0{,}05$ chấp nhận giả thiết H_0

- Nếu $P < 0{,}05$ bác bỏ giả thiết H_0 đồng nghĩa với việc chấp nhận H_1

3.11.5. Ví dụ

- **Ví dụ về bảng tương liên 2×2.** Hawkins và cộng sự (1993) đã tiến hành nghiêm cứu ảnh hưởng của việc thiến đến sự xuất hiện bệnh tiểu đường ở chuột. Biết rằng tỷ lệ mắc mắc bệnh này ở chuột là 24% đối với con đực và 73 % ở con cái. Từ 100 chuột thí nghiệm, chia một cách hoàn toàn ngẫu nhiên về 1 trong 2 cách sử lý thiến và không thiến. Số chuột ở 2 lô thí nghiệm được theo dõi cho đến 140 ngày tuổi và tiến hành lấy mẫu nghiên cứu từ 42 ngày tuổi. Bệnh tiểu đường được xác định đối với chuột có hàm lượng đường trong máu lớn hơn 200 mg/ dl. Kết quả thí nghiệm được ghi lại ở bảng sau:

	Cách xử lý		
	Thiến	Không thiến	Tổng
Mắc bệnh	26	12	38
Không mắc bệnh	24	38	62
Tổng số	50	50	100

Câu hỏi đặt ra: Tỷ lệ mắc bệnh ở 2 cách xử lý có khác nhau không?

Lời giải

Giả thiết H_0: $\pi_1 = \pi_2$,

H_1: $\pi_1 \neq \pi_2$,

Nhập tần suất quan sát (*O*) vào bảng tương liên 2×2

	Cách xử lý		
	Thiến	Không thiến	Tổng
Mắc bệnh	26	12	38
Không mắc bệnh	24	38	62
Tổng số	50	50	100

Ước tính tần suất theo lý thuyết (*E*):

	Cách xử lý		
Vacxin	Thiến	Không thiến	Tổng
Mắc bệnh	$=\frac{50\times38}{100}=19$	$=\frac{50\times38}{100}=19$	38
Không mắc bệnh	$=\frac{50\times62}{100}=31$	$=\frac{50\times62}{100}=31$	62
Tổng số	50	50	100

Tính giá trị χ^2 thực nghiệm theo công thức sau

$$\boxed{\chi^2 = \sum\frac{(|O-E|-0{,}5)^2}{E}}$$

$$\chi^2 = \frac{(|26-19|-0,5)^2}{19} + \cdots + \frac{(|38-31|-0,5)^2}{31} = 7,17$$

df = (Số hàng - 1) $\times$(Số cột – 1) = (2 - 1) $\times$ (2 - 1) = 1

Xác định giá trị *P* bằng cách so sánh giá trị χ^2 = 7,17 với df = 1

ta có 0,005 < P < 0,01

Rút ra kết luận

P< 0,05 bác bỏ giả thiết H_0; chứng tỏ rằng tỷ lệ chuột sau khi thiến mắc bệnh đái đường cao hơn so với chuột không bị thiến (π_1 = 26/50 = 0,52; π_2 = 12/50 = 0,24).

- **Ví dụ về bảng tương liên 4×3**. Đây là các loại vacxin phòng bệnh được so sánh với đối chứng. Với các mức độ: không, trung bình và nhiễm bệnh nặng được ghi lai sau 24 tháng. Số liệu được trình bày ở bảng sau:

Vacxin	Mức độ bệnh Không	 Trung bình	 Nặng	Tổng
Đối chứng	100 (137.3)	71 (42.6)	29 (20.1)	200
A	146 (133.9)	32 (41.6)	17 (19.6)	195
B	149 (132.5)	28 (41.2)	16 (19.3)	193
C	146 (137.3)	37 (42.6)	17 (20.1)	200
Tổng số	541	168	79	788

Bảng cho ta thấy tần suất *O* (*E*) - có nghĩa là tần suất quan sát và ước tính.

Chúng ta kiểm tra H_0 rằng không có mối liên quan nào giữa mức độ bệnh và tiêm phòng vacxin, nghĩa là tất cả các đều có cùng mức độ ảnh hưởng.

- Nếu H_0 đúng thì tần suất ước tính được tính như sau:

Nhóm đối chứng, không nhiễm bệnh (ô đầu tiên trong bảng):

E_1 = [(Tổng số hàng thứ nhất) $\times$ (Tổng số cột thứ nhất)] /Tổng số toàn bộ quan sát

= (200 $\times$ 541) /788 = 137,3

Chúng được biểu hiện trong dấu (....) của bảng. Tương tự như vậy ta có thể tính được các giá trị còn lại.

- Tính χ^2 thực nghiệm

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= 10.14 + 18.86 + 3.99 +$$
$$1.10 + 2.21 + 0.33 +$$
$$2.05 + 4.20 + 0.58 +$$
$$0.55 + 0.75 + 0.46$$
$$= 45.2$$

- Bậc tự do (*df*):

df = (Số hàng - 1) ×(Số cột - 1)

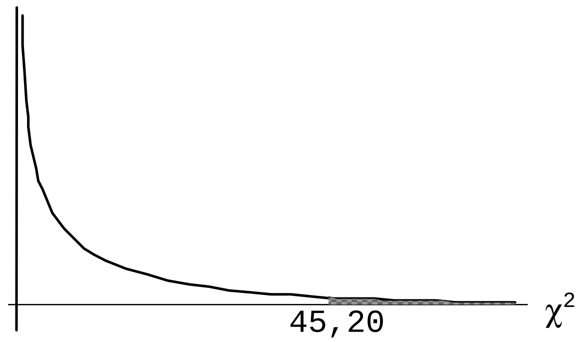
= (4 - 1) × (3-1) = 6

Xác định giá trị *P*

$$P = P(\chi^2_6 > 45.2) = 0.000$$

Kết luận:

Có mối liên hệ có ý nghĩa giữa các loại vacxin và mức độ nhiễm bệnh - có nghĩa là các mức độ khác nhau của nghiệm thức cho mức độ miễn dịch khác nhau (*P*=0,000). Kiểm tra lại tần suất *O* so với *E* ta thấy nhóm đối chứng có số trường hợp mắc bệnh cao hơn (không có vacxin).



Áp dụng Minitab

Chúng ta sẽ nhập số liệu vào WorkSheet của *Minitab* như sau:

Data Display

Row	Khong	TB	Nang
1	100	71	29
2	146	32	17
3	149	28	16
4	146	37	17

Lưu ý rằng ta phải ngầm hiểu rằng row1 = đối chứng, row2, row3, row4 tương ứng với các vacxin A, B, C.

Thực hiện phép thử bằng *Minitab*

Chi-Square Test: Khong, TB, Nang **Stat > Tables > Chi-Square Test…**

```
Expected counts are printed below observed counts

         Khong         TB      Nang    Total
    1      100         71        29      200
        137.31      42.64     20.05

    2      146         32        17      195
        133.88      41.57     19.55

    3      149         28        16      193
        132.50      41.15     19.35

    4      146         37        17      200
        137.31      42.64     20.05

Total      541        168        79      788

Chi-Sq = 10.138 + 18.863 +  3.994 +
          1.098 +  2.205 +  0.332 +
          2.054 +  4.201 +  0.580 +
          0.550 +  0.746 +  0.464 = 45.224
DF = 6, P-Value = 0.000
```

Ta cũng có kết quả tương tự như phần tính tay.

Lưu ý: Công thức ở trang 56 $\chi^2 = \sum \frac{\left(|O-E|-0,5\right)^2}{E}$ chỉ sử dụng đối với những thí nghiệm có số lượng quan sát hạn chế. Các phần mềm thống kê không thực hiện theo công thức này; vì vậy trước khi xử lý số liệu nên xem xét và quyết định ***có sử dụng được phần mềm thống kê hay không!!!***

3.12. **Bài kiểm tra số 4**

Khi tiến hành thử nghiệm 2 loại vacxin (A và B); đối với vacxin A, sau khi tiến hành thí nghiệm trên 1350 gà thí nghiệm quan sát thấy 250 con chết; tương tự với vacxin B, từ 2535 gà thấy 500 con chết. Một bác sỹ thú y quyết định chọn vacxin A đưa vào phòng bệnh cho gà ở trại chăn nuôi của mình. Anh (chị) cho biết quyết định đó đúng hay sai, vì sao?

4. Phụ lục

BẢNG SỐ NGẪU NHIÊN

81 37	66 40	77 65	29 99	77 42	92 78	15 25	07 76	79 24	21 84
48 03	48 91	03 57	56 56	42 76	57 27	60 60	16 30	76 96	94 49
86 49	52 63	66 70	80 71	09 64	84 36	03 54	53 39	36 30	69 27
73 59	16 61	43 18	86 80	19 42	23 78	86 08	44 08	55 51	12 97
10 46	82 01	40 55	50 91	24 12	34 43	20 37	71 52	13 25	67 31
63 34	98 49	54 23	60 36	10 40	08 12	34 46	59 82	91 74	60 92
18 40	40 07	42 21	10 22	39 57	86 80	03 29	64 96	73 84	72 47
59 86	66 45	91 17	29 15	92 05	97 60	76 48	44 58	89 64	01 26
30 99	69 70	16 08	76 29	74 90	18 42	43 71	47 22	10 21	08 69
14 49	02 64	25 44	27 12	36 82	67 84	58 21	61 72	45 23	63 43
99 76	35 87	72 35	14 61	70 33	94 30	18 23	70 30	80 72	72 04
50 42	77 64	94 44	17 80	67 98	72 15	00 52	41 76	16 85	33 23
10 38	18 55	57 31	38 12	97 80	91 47	94 45	67 92	31 55	16 91
46 52	61 13	33 04	30 47	97 11	30 03	87 98	33 06	29 77	56 41
29 21	02 78	61 84	33 50	43 75	42 28	40 16	12 42	03 44	10 28
83 59	26 14	81 77	04 94	98 12	33 71	07 29	35 25	86 82	52 43
87 22	31 54	76 04	80 79	92 37	97 31	53 34	10 57	19 48	32 86
73 53	23 83	40 45	57 33	18 29	13 61	64 03	38 09	01 88	13 14
29 32	83 46	27 05	18 31	46 93	59 83	90 79	53 91	47 02	26 90
70 71	37 04	12 71	30 23	31 51	92 96	09 93	08 52	94 79	45 34
87 29	28 54	53 54	33 39	22 61	46 98	84 24	28 71	42 75	98 07
83 78	88 92	75 35	07 41	70 05	83 13	45 06	24 89	75 66	06 27
69 26	97 35	72 95	58 30	84 12	70 41	36 92	05 62	89 01	62 31
07 82	88 94	99 80	07 37	94 52	15 26	90 39	39 51	53 40	98 78
55 80	29 81	32 27	28 59	29 74	27 46	15 47	00 47	94 04	03 43
80 73	03 69	35 68	22 77	82 26	83 58	62 71	77 88	00 70	45 58
45 69	97 79	98 33	45 64	83 62	20 36	34 64	67 29	08 47	56 72
25 15	57 13	07 95	01 02	02 70	86 74	56 14	94 33	49 73	62 71
82 87	56 32	99 86	35 13	22 12	25 90	89 20	82 87	46 23	14 27
00 98	13 94	00 85	09 30	97 98	72 40	81 87	33 96	58 28	08 64
61 99	16 38	11 08	28 65	70 71	79 51	31 38	27 99	64 57	99 98
79 93	50 34	41 50	21 49	74 52	03 52	53 24	89 53	96 19	31 06
36 19	99 62	65 08	46 68	44 96	73 98	65 41	72 37	46 27	11 41
88 27	35 22	39 59	19 39	65 55	59 20	25 48	23 61	78 35	48 89
24 20	27 94	31 17	47 50	37 11	15 19	46 34	23 80	37 60	30 50
54 55	44 08	73 05	63 52	47 43	82 40	98 97	92 13	46 31	02 67
83 93	99 35	06 85	63 39	04 12	93 91	86 88	63 68	62 75	91 38
64 64	87 77	53 05	29 76	06 23	88 81	10 33	02 86	86 93	12 00
74 72	31 23	20 17	06 56	26 91	86 60	48 28	08 93	56 03	26 44
81 76	68 15	22 70	38 56	71 59	69 38	45 64	79 98	69 02	11 90

BẢNG XÁC SUẤT CỦA PHÂN BỐ TIÊU CHUẨN HOÁ

Các giá trị trong bảng là của phân bố chuẩn với trung bình bằng 0 và độ lệch chuẩn là 1. Ứng với mỗi giá trị *z* trong là giá trị *P*, $P(Z < z)$.

z	*P*	*z*	*P*	*z*	*P*	*z*	*P*
-4,00	0,00003	-1,50	0,0668	0,00	0,5000	1,55	0,9394
-3,50	0,00023	-1,45	0,0735	0,05	0,5199	1,60	0,9452
-3,00	0,0013	-1,40	0,0808	0,10	0,5398	1,65	0,9505
-2,95	0,0016	-1,35	0,0885	0,15	0,5596	1,70	0,9554
-2,90	0,0019	-1,30	0,0968	0,20	0,5793	1,75	0,9599
-2,85	0,0022	-1,25	0,1056	0,25	0,5987	1,80	0,9641
-2,80	0,0026	-1,20	0,1151	0,30	0,6179	1,85	0,9678
-2,75	0,0030	-1,15	0,1251	0,35	0,6368	1,90	0,9713
-2,70	0,0035	-1,10	0,1357	0,40	0,6554	1,95	0,9744
-2,65	0,0040	-1,05	0,1469	0,45	0,6736	2,00	0,9772
-2,60	0,0047	-1,00	0,1587	0,50	0,6915	2,05	0,9798
-2,55	0,0054	-0,95	0,1711	0,55	0,7088	2,10	0,9821
-2,50	0,0062	-0,90	0,1841	0,60	0,7257	2,15	0,9842
-2,45	0,0071	-0,85	0,1977	0,65	0,7422	2,20	0,9861
-2,40	0,0082	-0,80	0,2119	0,70	0,7580	2,25	0,9878
-2,35	0,0094	-0,75	0,2266	0,75	0,7734	2,30	0,9893
-2,30	0,0107	-0,70	0,2420	0,80	0,7881	2,35	0,9906
-2,25	0,0122	-0,65	0,2578	0,85	0,8023	2,40	0,9918
-2,20	0,0139	-0,60	0,2743	0,90	0,8159	2,45	0,9929
-2,15	0,0158	-0,55	0,2912	0,95	0,8289	2,50	0,9938
-2,10	0,0179	-0,50	0,3085	1,00	0,8413	2,55	0,9946
-2,05	0,0202	-0,45	0,3264	1,05	0,8531	2,60	0,9953
-2,00	0,0228	-0,40	0,3446	1,10	0,8643	2,65	0,9960
-1,95	0,0256	-0,35	0,3632	1,15	0,8749	2,70	0,9965
-1,90	0,0287	-0,30	0,3821	1,20	0,8849	2,75	0,9970
-1,85	0,0322	-0,25	0,4013	1,25	0,8944	2,80	0,9974
-1,80	0,0359	-0,20	0,4207	1,30	0,9032	2,85	0,9978
-1,75	0,0401	-0,15	0,4404	1,35	0,9115	2,90	0,9981
-1,70	0,0446	-0,10	0,4602	1,40	0,9192	2,95	0,9984
-1,65	0,0495	-0,05	0,4801	1,45	0,9265	3,00	0,9987
-1,60	0,0548	0,00	0,5000	1,50	0,9332	3,50	0,99977
-1,55	0,0606					4,00	0,99997

Một vài giá trị tới hạn của *z*:

P	0,80	0,90	0,95	0,975	0,99	0,995	0,999
z	0,842	1,282	1,645	1,960	2,326	2,576	3,090

BẢNG XÁC SUẤT CỦA PHÂN BỐ STUDENT (T)

Các giá trị trong bảng là của phân bố *t*. Cột thứ nhất là bậc tự do (df). Các cột còn lại cho ta các giá trị lý thuyết về kiểm định một hướng (phần trên); $P(T_{df} > t) = P$, hoặc 2 hướng; $P(T_{df} > t$ hoặc $T_{df} < -t) = P$ trong đó P là mức xác suất được thể hiện ở đầu cột.

df	*P*						
	0,10	0,05	0,025	0,01	0,005	0,001	(1 hướng)
	0,20	0,10	0,05	0,02	0,01	0,002	(2 hướng)
1	3,078	6,314	12,706	31,821	63,657	318,313	
2	1,886	2,920	4,303	6,965	9,925	22,327	
3	1,638	2,353	3,182	4,541	5,841	10,215	
4	1,533	2,132	2,776	3,747	4,604	7,173	
5	1,476	2,015	2,571	3,365	4,032	5,893	
6	1,440	1,943	2,447	3,143	3,707	5,208	
7	1,415	1,895	2,365	2,998	3,499	4,785	
8	1,397	1,860	2,306	2,896	3,355	4,501	
9	1,383	1,833	2,262	2,821	3,250	4,297	
10	1,372	1,812	2,228	2,764	3,169	4,144	
11	1,363	1,796	2,201	2,718	3,106	4,025	
12	1,356	1,782	2,179	2,681	3,055	3,930	
13	1,350	1,771	2,160	2,650	3,012	3,852	
14	1,345	1,761	2,145	2,624	2,977	3,787	
15	1,341	1,753	2,131	2,602	2,947	3,733	
16	1,337	1,746	2,120	2,583	2,921	3,686	
17	1,333	1,740	2,110	2,567	2,898	3,646	
18	1,330	1,734	2,101	2,552	2,878	3,611	
19	1,328	1,729	2,093	2,539	2,861	3,579	
20	1,325	1,725	2,086	2,528	2,845	3,552	
21	1,323	1,721	2,080	2,518	2,831	3,527	
22	1,321	1,717	2,074	2,508	2,819	3,505	
23	1,319	1,714	2,069	2,500	2,807	3,485	
24	1,318	1,711	2,064	2,492	2,797	3,467	
25	1,316	1,708	2,060	2,485	2,787	3,450	
26	1,315	1,706	2,056	2,479	2,779	3,435	
27	1,314	1,703	2,052	2,473	2,771	3,421	
28	1,313	1,701	2,048	2,467	2,763	3,408	
29	1,311	1,699	2,045	2,462	2,756	3,396	
30	1,310	1,697	2,042	2,457	2,750	3,385	
40	1,303	1,684	2,021	2,423	2,704	3,307	
60	1,296	1,671	2,000	2,390	2,660	3,232	
120	1,289	1,658	1,980	2,358	2,617	3,160	
∞	1,282	1,645	1,960	2,326	2,576	3,090	

BẢNG XÁC SUẤT CỦA PHÂN BỐ KHI BÌNH PHƯƠNG (χ^2)

Giá trị trong bảng là của phân bố χ^2. Cột thứ nhất là bậc tự do (df). Các cột còn lại cho ta các giá trị lý thuyết ở phần đuôi; $P(\chi^2_{df} > x^2) = P$, trong đó P là mức xác suất thể hiện ở đầu cột.

df	*P*					
	0,10	0,05	0,025	0,01	0,005	0,001
1	2,71	3,84	5,02	6,63	7,88	10,83
2	4,61	5,99	7,38	9,21	10,60	13,82
3	6,25	7,81	9,35	11,34	12,84	16,27
4	7,78	9,49	11,14	13,28	14,86	18,47
5	9,24	11,07	12,83	15,09	16,75	20,51
6	10,64	12,59	14,45	16,81	18,55	22,46
7	12,02	14,07	16,01	18,48	20,28	24,32
8	13,36	15,51	17,53	20,09	21,95	26,12
9	14,68	16,92	19,02	21,67	23,59	27,88
10	15,99	18,31	20,48	23,21	25,19	29,59
11	17,28	19,68	21,92	24,73	26,76	31,26
12	18,55	21,03	23,34	26,22	28,30	32,91
13	19,81	22,36	24,74	27,69	29,82	34,53
14	21,06	23,68	26,12	29,14	31,32	36,12
15	22,31	25,00	27,49	30,58	32,80	37,70
16	23,54	26,30	28,85	32,00	34,27	39,25
17	24,77	27,59	30,19	33,41	35,72	40,79
18	25,99	28,87	31,53	34,81	37,16	42,31
19	27,20	30,14	32,85	36,19	38,58	43,82
20	28,41	31,41	34,17	37,57	40,00	45,31
21	29,62	32,67	35,48	38,93	41,40	46,80
22	30,81	33,92	36,78	40,29	42,80	48,27
23	32,01	35,17	38,08	41,64	44,18	49,73
24	33,20	36,42	39,36	42,98	45,56	51,18
25	34,38	37,65	40,65	44,31	46,93	52,62
26	35,56	38,89	41,92	45,64	48,29	54,05
27	36,74	40,11	43,19	46,96	49,65	55,48
28	37,92	41,34	44,46	48,28	50,99	56,89
29	39,09	42,56	45,72	49,59	52,34	58,30
30	40,26	43,77	46,98	50,89	53,67	59,70
40	51,81	55,76	59,34	63,69	66,77	73,40
50	63,17	67,50	71,42	76,15	79,49	86,66
60	74,40	79,08	83,30	88,38	91,95	99,61
80	96,58	101,88	106,63	112,33	116,32	124,84
100	118,50	124,34	129,56	135,81	140,17	149,45

Đối với trường hợp bậc tự do lớn ta có thể tính toán như sau, áp dụng phân bố chuẩn cho χ^2, $z = \sqrt{2\chi^2} - \sqrt{2 \times \mathrm{df} - 1}$, và so sánh giá trị z với "Bảng xác suất của phân bố tiêu chuẩn hoá"

BẢNG XÁC SUẤT CỦA PHÂN BỐ FISHER

Trong bảng là giá trị của phân bố Fisher *F*. Bậc tự do (ν_1) xác định vị trí của cột và bậc tự do (ν_2) xác định vị trí của hàng. Các giá trị trong bảng là giá trị lý thuyết của phần đuôi trên; $P = (F_{\nu 1,\, \nu 2} > f) = P$, trong đó *P* là xác suất (0,10; 0,05; 0,01).

											ν_1										
ν_2	*P*	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120	∞
1	0,10	39,86	49,50	53,59	55,83	57.24	58.20	58.91	59.44	59.86	60.19	60.47	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
	0,05	161,4	199,5	215,7	224,6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
	0,01	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6083	6107	6157	6209	6234	6260	6286	6313	6340	6366
2	0,10	8,53	9,00	9,16	9,24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
	0,05	18,51	19,00	19,16	19,25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
	0,01	98,50	99,00	99,16	99,25	99.30	99.33	99.36	99.38	99.39	99.40	99.41	99.42	99.43	99.45	99.46	99.47	99.48	99.48	99.49	99.50
3	0,10	5,54	5,46	5,39	5,34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
	0,05	10,13	9,55	9,28	9,12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
	0,01	34,12	30,82	29,46	28,71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	0,10	4,54	4,32	4,19	4,11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
	0,05	7,71	6,94	6,59	6,39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
	0,01	21,20	18,00	16,69	15,98	15.52	15.21	14.98	14.80	14.66	14.55	14.45	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	0,10	4,06	3,78	3,62	3,52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
	0,05	6,61	5,79	5,41	5,19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
	0,01	16,26	13,27	12,06	11,39	10.97	10.67	10.46	10.29	10.16	10.05	9.96	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	0,10	3,78	3,46	3,29	3,18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
	0,05	5,99	5,14	4,76	4,53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
	0,01	13,75	10,92	9,78	9,15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	0,10	3,59	3,26	3,07	2,96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
	0,05	5,59	4,74	4,35	4,12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
	0,01	12,25	9,55	8,45	7,85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	0,10	3,46	3,11	2,92	2,81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
	0,05	5,32	4,46	4,07	3,84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
	0,01	11,26	8,65	7,59	7,01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	0,10	3,36	3,01	2,81	2,69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
	0,05	5,12	4,26	3,86	3,63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
	0,01	10,56	8,02	6,99	6,42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31

ν_2	*P*	ν_1: 1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120	∞
10	0,10	3,29	2,92	2,73	2,61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
	0,05	4,96	4,10	3,71	3,48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
	0,01	10,04	7,56	6,55	5,99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	0,10	3,23	2,86	2,66	2,54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
	0,05	4,84	3,98	3,59	3,36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
	0,01	9,65	7,21	6,22	5,67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	0,10	3,18	2,81	2,61	2,48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
	0,05	4,75	3,89	3,49	3,26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
	0,01	9,33	6,93	5,95	5,41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
15	0,10	3,07	2,70	2,49	2,36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
	0,05	4,54	3,68	3,29	3,06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
	0,01	8,68	6,36	5,42	4,89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
20	0,10	2,97	2,59	2,38	2,25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
	0,05	4,35	3,49	3,10	2,87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
	0,01	8,10	5,85	4,94	4,43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
24	0,10	2,93	2,54	2,33	2,19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
	0,05	4,26	3,40	3,01	2,78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
	0,01	7,82	5,61	4,72	4,22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
30	0,10	2,88	2,49	2,28	2,14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
	0,05	4,17	3,32	2,92	2,69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
	0,01	7,56	5,39	4,51	4,02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	0,10	2,84	2,44	2,23	2,09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
	0,05	4,08	3,23	2,84	2,61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
	0,01	7,31	5,18	4,31	3,83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	0,10	2,79	2,39	2,18	2,04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
	0,05	4,00	3,15	2,76	2,53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
	0,01	7,08	4,98	4,13	3,65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	0,10	2,75	2,35	2,13	1,99	1.90	1.82	1.77	1.72	1.68	1.65	1.63	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
	0,05	3,92	3,07	2,68	2,45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
	0,01	6,85	4,79	3,95	3,48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	0,10	2,71	2,30	2,08	1,94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00
	0,05	3,84	3,00	2,60	2,37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00
	0,01	6,63	4,61	3,78	3,32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

BẢNG GIÁ TRỊ 2½% PHÍA TRÊN CỦA PHÂN BỐ FISHER F

Giá trị trong bảng là của phân bố Fisher *F*. Bậc tự do (ν_1) xác định vị trí của cột và bậc tự do (ν_2) xác định vị trí của hàng. Các giá trị trong bảng là giá trị lý thuyết tại điểm 2,5%; $P(F_{\nu_1,\nu_2} > f) = 0{,}025$.

														ν_1					
ν_2	1	2	3	4	5	6	7	8	9	10	11	12	15	20	24	30	40	60	120	∞
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,6	963,3	968,6	973,0	976,7	984,9	993,1	997,3	1001	1006	1010	1014	1018
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	17,44	16,04	15,44	15,10	14,88	14,73	14,62	14,54	14,47	14,42	14,37	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,79	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,57	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46	5,41	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76	4,71	4,67	4,57	4,47	4,41	4,36	4,31	4,25	4,20	4,14
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,24	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,91	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,66	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,47	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,32	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	3,01	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,72	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,59	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,46	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,33	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,22	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,10	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,99	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00

5. Tài liệu tham khảo

5.1. Tiếng Việt

- Pascal Leroy, Frederic Farnir (1999). *Thống kê sinh học*. Tài liệu dịch từ nguyên bản tiếng Pháp; người dịch Đặng Vũ Bình. Đại học Nông nghiệp I Hà Nội.
- Phạm Chí Thành (1988). *Phương pháp thí nghiệm đồng ruộng*. Đại học Nông nghiệp I Hà Nội.
- Phan Hiếu Hiền (2001). *Phương pháp bố trí thí nghiệm*. Nhà xuất bản Nông Nghiệp.
- Chu Văn Mẫn, Đào Hữu Hồ (1999). *Thống kê sinh học*. Nhà xuất bản Khoa học và kỹ thuật.
- Nguyễn Văn Thiện (1997). *Phương pháp nghiên cứu trong chăn nuôi*. Nhà xuất bản Nông nghiệp.

5.2. Tiếng Anh

- R.C. Campbell (2000). *Statistics for Biologists*. Cambridge University Press.
- Aviva Petrie and Paul Watson (2001). *Statistics for veterinary and animal science.* Blackwell Science.
- R. Mead, R.N. Curnow and A.M. Hasted (1993). *Statistical methods in agriculture and experimental biology*. Chapman & Hall/Crc.
- W.G. Cochran and G.M. Cox (1966). *Experimental Designs*. Wiley International Edition.
- D.R.Cox (1958). *Planning of experiments*. Wiley International Edition.
- Robert R. Sokal, F. James Rohlf (2000). *Biometry*. W.H. Freeman and Company.
- Mick O'Neill, Peter Thomson (2002). *Third year biometry: Experimental design, Statistical modelling*. The University of Sydney.
- Peter Thomson, Frank Nicholas, Cris Moran (2002). *Genetics and biometry*. The University of Sydney.
- Douglas C. Montgomery (1996). *Design and analysis of experiments*. Wiley International Edition.
- Harold R. Lindman (1991). *Analysis of variance in experimental design*. Springer-Verlag.
- *Meet Minitab*, release 13 for Windows®. Minitab Inc.
- *Minitab user's guide 1*, release 13 for Windows®. Minitab Inc.
- *Minitab user's guide 2*, release 13 for Windows®. Minitab Inc.

5.3. Tiếng Nga

- Б.А. Доспехов (1985). *Методика полевого опыта*. Агропромиздат.
- А.И. Овсянников (1976). *Основы опытного дела в животноводстве*. Колос.

5.4. Tiếng Pháp

- Claustriaux J.J. (2002). *Expérimentation, concevoir pour analyser*. Gembloux, faculté universitaire des sciences agronomique.