

HỌC VIỆN NÔNG NGHIỆP VIỆT NAM

NGUYỄN XUÂN TRẠCH & ĐỖ ĐỨC LỰC

GIÁO TRÌNH
PHÂN TÍCH SỐ LIỆU THÍ NGHIỆM VÀ
CÔNG BỐ KẾT QUẢ NGHIÊN CỨU CHĂN NUÔI

(Dành cho nghiên cứu sinh ngành Chăn nuôi)

NHÀ XUẤT BẢN ĐẠI HỌC NÔNG NGHIỆP – 2016

LỜI NÓI ĐẦU

Mục tiêu đào tạo tiến sĩ là đào tạo những người làm nghiên cứu khoa học trong một lĩnh vực nhất định, có khả năng làm việc độc lập, biết tiếp thu kiến thức của nhân loại, biết nghiên cứu khoa học và truyền tải kiến thức cũng như kết quả nghiên cứu cho người khác. Nếu coi việc nhận bằng Tiến sĩ là đánh dấu cho sự bắt đầu của một sự nghiệp khoa học thì làm nghiên cứu sinh là một giai đoạn chuẩn bị cho sự bắt đầu đó. Trong giai đoạn tập dượt làm khoa học này, việc học tập phương pháp là quan trọng nhất.

Giáo trình được biên soạn nhằm giúp nghiên cứu sinh phát triển các kỹ năng cần thiết để làm hành trang trên con đường khoa học “từ ý tưởng đến nghiên cứu” và “từ nghiên cứu đến công bố”. Đó là kỹ năng xây dựng ý tưởng nghiên cứu, lập kế hoạch, triển khai nghiên cứu, phân tích số liệu thí nghiệm và công bố kết quả nghiên cứu. Trọng tâm của Giáo trình này là phương pháp sử dụng phần mềm thống kê để phân tích số liệu cho các mô hình thiết kế thí nghiệm thường gặp trong nghiên cứu chăn nuôi, chủ yếu là các mô hình phân tích phương sai, phân tích hồi quy và mô hình tuyến tính tổng quát. Hơn nữa, một phần quan trọng của Giáo trình dành cho việc luyện kỹ năng viết và thuyết trình, từ viết đề cương nghiên cứu, thuyết trình báo cáo khoa học, đến viết bài báo khoa học và luận án Tiến sĩ.

Các đối tượng khác như sinh viên đại học và học viên cao học cũng có thể tham khảo để phục vụ học tập và phát triển kỹ năng nghiên cứu khoa học. Nghiên cứu sinh các ngành khác hay các cán bộ nghiên cứu và cán bộ giảng dạy đại học trong khối Nông-Lâm-Ngư cũng có thể tham khảo tài liệu này để phát triển kỹ năng phân tích thống kê và viết tài liệu khoa học.

Tham gia biên soạn Giáo trình này gồm PGS. TS. Nguyễn Xuân Trạch và TS. Đỗ Đức Lực, giảng viên Khoa Chăn nuôi, Học viện Nông nghiệp Việt Nam. Nhóm tác giả xin chân thành cảm ơn Nhà giáo ưu tú Nguyễn Đình Hiền, GS. TS. Vũ Chí Cương, PGS. TS. Vũ Đình Tôn, PGS. TS. Vũ Đình Hòa, TS. Nguyễn Văn Hạnh đã có nhiều ý kiến đóng góp giúp hoàn thiện giáo trình này.

Mặc dù đã có nhiều cố gắng trong quá trình biên soạn, song không thể tránh được những thiếu sót, các tác giả rất mong nhận được sự góp ý của bạn đọc để lần xuất bản sau được hoàn thiện hơn.

CÁC TÁC GIẢ

MỤC LỤC

Chương 1. MỘT SỐ VẤN ĐỀ CƠ BẢN VỀ NGHIÊN CỨU CHĂN NUÔI.....	1
1.1. KHÁI NIỆM VỀ NGHIÊN CỨU KHOA HỌC	1
1.1.1. Tri thức khoa học.....	1
1.1.2. Mục tiêu và bản chất của nghiên cứu khoa học.....	2
1.1.3. Phương pháp nghiên cứu khoa học	2
1.2. PHƯƠNG PHÁP NGHIÊN CỨU CHĂN NUÔI	3
1.2.1. Câu hỏi nghiên cứu.....	4
1.2.2. Cơ sở lý luận và thực tiễn (Tổng quan tài liệu)	5
1.2.3. Giả thuyết khoa học.....	7
1.2.4. Thí nghiệm chăn nuôi.....	8
1.2.5. Phân tích số liệu thí nghiệm	18
1.3. ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC	21
1.3.1. Đề cương nghiên cứu là gì?.....	21
1.3.2. Nội dung của đề cương nghiên cứu	22
1.3.3. Chi tiết các phần của đề cương nghiên cứu.....	23
CÂU HỎI ÔN TẬP	28
Chương 2. PHÂN TÍCH SỐ LIỆU THÍ NGHIỆM BẰNG PHẦN MỀM MÁY TÍNH.....	29
2.1. GIỚI THIỆU PHẦN MỀM THỐNG KÊ.....	29
2.1.1. Khởi động phần mềm	29
2.1.2. Nhập dữ liệu	30
2.1.3. Tạo câu lệnh	31
2.1.4. Kết quả phân tích.....	33
2.1.5. Chuyển kết quả từ phần mềm SAS qua một định dạng khác	34
2.2. PHÂN TÍCH PHƯƠNG SAI (ANOVA)	35
2.2.1. Kiểm định điều kiện phân tích phương sai.....	35
2.2.2. So sánh cặp đôi các giá trị trung bình.....	38
2.2.3. Các mô hình phân tích phương sai và so sánh cặp đôi	41
2.2.4. Phân tích hiệp phương sai (ANCOVA).....	81
2.3. TƯƠNG QUAN VÀ PHÂN TÍCH HỒI QUY	85
2.3.1. Tương quan	85
2.3.2. Hồi quy tuyến tính đơn và đa biến	87
2.3.3. Lựa chọn mô hình hồi quy tốt nhất.....	91

2.3.4. Hồi quy đa thức	95
2.3.5. Hồi quy phi tuyến	98
2.4. PHÂN TÍCH SỐ LIỆU ĐỊNH TÍNH	102
2.4.1. Phép thử χ^2 và Fisher exact test.....	102
2.4.2. Hồi quy logistic	106
2.5. BIẾN ĐỔI SỐ LIỆU	109
2.6. PHÂN TÍCH PHƯƠNG SAI PHI THAM SỐ	115
CÂU HỎI ÔN TẬP	118
Chương 3. CÔNG BỐ KẾT QUẢ NGHIÊN CỨU	120
3.1. TRÌNH BÀY KẾT QUẢ NGHIÊN CỨU.....	120
3.1.1. Bảng số liệu.....	120
3.1.2. Biểu đồ	124
3.1.3. Văn viết mô tả số liệu.....	128
3.2. BÁO CÁO KHOA HỌC.....	129
3.2.1. Chuẩn bị báo cáo	129
3.2.2. Thuyết trình báo cáo.....	133
3.2.3. Trả lời chất vấn sau thuyết trình.....	136
3.3. BÀI BÁO KHOA HỌC	137
3.3.1. Cấu trúc bài báo khoa học	137
3.3.2. Chuẩn bị trước khi viết.....	138
3.3.3. Cách viết các phần của bài báo.....	140
3.4. LUẬN ÁN TIẾN SĨ	149
3.4.1. Yêu cầu chung đối với luận án tiến sĩ	149
3.4.2. Cấu trúc của luận án	150
3.4.3. Nội dung và cách viết các phần chính của luận án.....	151
3.5. VĂN PHONG VÀ ĐẠO VĂN TRONG CÔNG BỐ KHOA HỌC.....	155
3.5.1. Văn phong khoa học.....	155
3.5.2. Đạo văn trong công bố khoa học.....	158
CÂU HỎI ÔN TẬP	159
Chương 4. THỰC HÀNH PHÂN TÍCH VÀ CÔNG BỐ KẾT QUẢ THÍ NGHIỆM	160
4.1. NỘI DUNG THỰC HÀNH	160
4.2. BÀI THỰC HÀNH	160
4.2.1. Thí nghiệm	160
4.2.2. Yêu cầu thực hành.....	162
TÀI LIỆU THAM KHẢO	163

Chương 1

MỘT SỐ VẤN ĐỀ CƠ BẢN VỀ NGHIÊN CỨU CHĂN NUÔI

Trong nghiên cứu chăn nuôi, cũng như các lĩnh vực nghiên cứu khác, một nghiên cứu khoa học được bắt đầu từ ý tưởng nghiên cứu về một vấn đề nào đó, từ đó đặt ra câu hỏi cho việc nghiên cứu nhằm có được kết quả hữu ích lấp vào khoảng trống tri thức hiện tại hay giải quyết những vấn đề nảy sinh từ thực tiễn. Nhưng “từ ý tưởng đến nghiên cứu”, nhà khoa học có rất nhiều việc phải làm. Trước tiên, ý tưởng đó phải được chuyển thành một thiết kế nghiên cứu hoàn chỉnh thể hiện dưới dạng một đề cương nghiên cứu được chấp nhận. Công việc tiếp theo là tiến hành thí nghiệm, thu thập số liệu, phân tích, đánh giá và công bố kết quả thu được. Chương đầu tiên này nhằm giúp nghiên cứu sinh hệ thống khái quát lại một số vấn đề cơ bản đó trong nghiên cứu chăn nuôi.

1.1. KHÁI NIỆM VỀ NGHIÊN CỨU KHOA HỌC

1.1.1. Tri thức khoa học

Tri thức (hay *kiến thức*) của nhân loại được hình thành trong lịch sử và không ngừng phát triển theo thời gian. Tri thức là một phạm trù rất rộng, có thể xem xét ở nhiều cấp độ, khía cạnh khác nhau. Tri thức có thể phân ra tri thức đời thường (còn gọi là tri thức tiền khoa học, tri thức kinh nghiệm đời thường hay tri thức thường nghiệm), tri thức nghệ thuật và tri thức khoa học. *Tri thức khoa học* là những hiểu biết được tích lũy một cách có hệ thống nhờ hoạt động *nghiên cứu khoa học*. Bản thân tri thức khoa học cũng bao gồm cả *tri thức thực nghiệm* và *tri thức lý luận*. Trong đó, tri thức thực nghiệm ở trình độ thấp, còn tri thức lý luận ở trình độ cao.

Tri thức kinh nghiệm đời thường là những hiểu biết được tích lũy qua hoạt động sống hàng ngày trong mối quan hệ giữa con người với con người và giữa con người với thiên nhiên. Quá trình này giúp con người hiểu biết về sự vật, về cách quản lý thiên nhiên và hình thành mối quan hệ giữa con người trong xã hội. Tri thức kinh nghiệm được con người không ngừng sử dụng và phát triển trong hoạt động thực tiễn. Tuy nhiên, tri thức kinh nghiệm chưa thật sự đi sâu vào bản chất, chưa thấy được hết các thuộc tính của sự vật và mối quan hệ bên trong giữa sự vật và con người. Vì vậy, tri thức kinh nghiệm chỉ phát triển đến một hiểu biết giới hạn nhất định.

Tri thức thực nghiệm khoa học chủ yếu thu nhận được thông qua quan sát và thử nghiệm khoa học. Nó mới phản ánh cái hiện tượng, cái cụ thể, cái trực tiếp, bề ngoài của sự vật... Để nắm bắt được bản chất sự vật thì nhận thức của con người tất yếu phải chuyển lên trình độ tri thức lý luận.

Tri thức lý luận khoa học được hình thành trên cơ sở khái quát hóa tri thức kinh nghiệm và thực nghiệm. Nó tồn tại trong hệ thống các khái niệm, phạm trù, quy luật, lý

thuyết, học thuyết nào đó. Tri thức lý luận ở vào trình độ cao nhất của tri thức khoa học,... là kết quả của quá trình nghiên cứu, học tập nghiêm túc, bền bỉ, có hệ thống của con người... là sự phản ánh sâu sắc hơn, đầy đủ hơn, chính xác hơn mối quan hệ giữa các sự vật và hiện tượng. Tri thức lý luận mang lại những hiểu biết có tính bản chất, bên trong, vạch ra những mối liên hệ tất nhiên và tính quy luật của đối tượng... cho phép con người tiến gần sát đến chân lý.

1.1.2. Mục tiêu và bản chất của nghiên cứu khoa học

Có nhiều giải nghĩa khác nhau về mục tiêu và bản chất của nghiên cứu khoa học. Chẳng hạn, “nghiên cứu khoa học là một giải pháp có hệ thống nhằm tìm lời giải cho các câu hỏi” (Tuckman, 1999). Nghiên cứu khoa học là một quá trình tìm hiểu tích cực, cần mẫn và có hệ thống nhằm phát hiện, diễn giải hay tổng hợp các bằng chứng, sự kiện, hành vi hay lý thuyết, hoặc là để ứng dụng vào thực tiễn nhờ những bằng chứng, định luật hay lý thuyết như vậy. Thuật ngữ “nghiên cứu” cũng được dùng để mô tả tập hợp các thông tin về một đối tượng cụ thể (*Encyclopedia Wikipedia*).

Như vậy, nghiên cứu khoa học là một hoạt động tìm kiếm, xem xét, điều tra, hoặc thử nghiệm. Dựa trên những số liệu, tài liệu, kiến thức,... có được từ các thí nghiệm nghiên cứu khoa học để phát hiện ra những cái mới về bản chất sự vật, về thế giới tự nhiên và xã hội. Những hoạt động tìm kiếm, xem xét, điều tra, hoặc thử nghiệm có hệ thống đó hướng tới kiến thức khoa học hoàn chỉnh hơn hay hiểu biết tốt hơn về sự vật nghiên cứu. Dựa trên những kiến thức mới đạt được từ các công trình nghiên cứu khoa học, các nhà khoa học có thể sáng tạo ra những phương pháp, sản phẩm mới, công nghệ mới có giá trị hơn.

1.1.3. Phương pháp nghiên cứu khoa học

Nghiên cứu khoa học đòi hỏi không những phải có mục tiêu rõ ràng mà phải sử dụng đúng phương pháp nghiên cứu khoa học. Đó là khái quát cách tiến hành các nghiên cứu khoa học, có thể coi như một chỉnh thể các nguyên tắc và các phương pháp giúp nhà nghiên cứu trong tất cả các lĩnh vực khoa học thu được các kết quả nghiên cứu xác thực, bao gồm những hướng dẫn rõ ràng và được công nhận rộng rãi để thu thập, đánh giá và chuyển tải thông tin trong bối cảnh của một nghiên cứu.

Phương pháp nghiên cứu khoa học bao gồm việc đưa ra *luận đề* (cần chứng minh điều gì?) và chứng minh nó thông qua *luận cứ* (chứng minh bằng cái gì?) và *luận chứng* (chứng minh bằng cách nào?). Nói một cách khác, phương pháp nghiên cứu khoa học đòi hỏi phải chọn phương pháp thích hợp (*luận chứng*) để chứng minh một phán đoán hay giả thuyết khoa học (*luận đề*) thông qua thu thập và xử lý thông tin (*luận cứ*). Việc lựa chọn phương pháp thích hợp cho một nghiên cứu khoa học thuộc phạm trù *phương pháp luận nghiên cứu*. Đó là suy luận logic của việc quyết định lựa chọn các phương pháp hay kỹ thuật được sử dụng trong nghiên cứu. Thông qua đó, người nghiên cứu có thể giải thích tại sao sử dụng phương pháp này mà không sử dụng phương pháp khác.

Các ngành khoa học khác nhau có thể có những phương pháp nghiên cứu khoa học khác nhau. Các ngành khoa học tự nhiên như vật lý, hoá học, nông nghiệp thường sử dụng *phương pháp nghiên cứu thực nghiệm*, như tiến hành bố trí thí nghiệm để thu thập số liệu, để giải thích và kết luận. Các ngành khoa học xã hội như nhân chủng học, kinh tế, lịch sử... thường sử dụng *phương pháp nghiên cứu phi thực nghiệm* thông qua thu thập thông tin từ điều tra, quan sát hay phỏng vấn. Phương pháp nghiên cứu khoa học cũng có những bước chung như: quan sát sự vật hay hiện tượng, đặt câu hỏi nghiên cứu, lập giả thuyết, thu thập và phân tích số liệu, thảo luận và rút ra kết luận. Tuy nhiên, vẫn có sự khác nhau về quá trình thu thập, xử lý và phân tích số liệu.

Nhìn chung, một nghiên cứu khoa học được bắt đầu từ *ý tưởng nghiên cứu* về một *vấn đề* nào đó. Vấn đề nghiên cứu là những điều chưa biết hoặc chưa biết thấu đáo về bản chất hoặc hiện tượng, cần được làm rõ trong quá trình nghiên cứu. Vì vậy, khi đã phát hiện được một vấn đề trong nghiên cứu, người nghiên cứu tất yếu sẽ nảy sinh hàng loạt ý tưởng giải quyết vấn đề. Ý tưởng đó được gọi là *ý tưởng nghiên cứu* (ý tưởng khoa học). Đây chính là cơ sở ban đầu để đi đến cách giải quyết vấn đề nghiên cứu.

Nhà khoa học phải thể hiện rõ mình muốn biết gì về vấn đề đó, tức là phải đặt ra được *câu hỏi nghiên cứu* (chưa có lời giải). Vấn đề và câu hỏi nghiên cứu đặt ra phải được chứng minh là có ý nghĩa về mặt khoa học và/hay thực tiễn trên cơ sở một tổng quan tài liệu (cơ sở lý luận và thực tiễn) cập nhật các thông tin liên quan cho thấy việc thực hiện ý tưởng nghiên cứu sẽ góp phần lấp khoảng trống trong tri thức hiện tại, hay nói một cách khác là công trình nghiên cứu có tính cấp thiết và sẽ bổ sung kiến thức mới.

Muốn trả lời câu hỏi nghiên cứu nhà khoa học phải đưa ra giả thuyết khoa học (câu trả lời giả định) để làm cơ sở cho việc thiết kế thí nghiệm nghiên cứu. Công việc tiếp theo là tổ chức điều tra hay làm thí nghiệm để thu thập số liệu khách quan. Sau khi thu thập được số liệu, nhà khoa học phải tiến hành xử lý, phân tích và thảo luận kết quả đó để đánh giá xem kết quả thu được nói lên điều gì? Đây là những phát hiện mới? Kết quả ủng hộ hay bác bỏ giả thuyết đã nêu ra? Ý nghĩa khoa học và thực tiễn của chúng? Có câu hỏi/giả thuyết nào mới xuất hiện tiếp không?

Cuối cùng, kết quả nghiên cứu phải được công bố dưới dạng báo cáo tại hội thảo khoa học, bài báo đăng trên tạp chí chuyên ngành, hay đối với học viên cao học và nghiên cứu sinh là luận văn/luận án để bảo vệ trước Hội đồng đánh giá.

1.2. PHƯƠNG PHÁP NGHIÊN CỨU CHĂN NUÔI

Việc tiến hành nghiên cứu khoa học phải thực hiện theo *quy trình nghiên cứu khoa học*. Tương tự như nghiên cứu trong nhiều lĩnh vực khoa học khác, quy trình nghiên cứu chăn nuôi có thể tóm tắt thành 7 bước như trong hình 1.1.



Hình 1.1. Các bước trong quy trình nghiên cứu chăn nuôi

1.2.1. Câu hỏi nghiên cứu

Câu hỏi nghiên cứu là một loại câu hỏi mang tính chất bất định về một “vấn đề” khoa học nào đó. Câu hỏi nghiên cứu là câu hỏi chưa có đáp án mà phải thông qua nghiên cứu khoa học thì mới có thể tìm được câu trả lời chính xác. *Đặt câu hỏi nghiên cứu là bước quan trọng nhất trong quy trình nghiên cứu khoa học* để trên cơ sở đó nhà khoa học đưa ra “giả thuyết” khoa học, rồi tiến hành thiết kế và tổ chức thí nghiệm để kiểm định giả thuyết đó. Chính vì thế, yêu cầu của câu hỏi nghiên cứu là phải đảm bảo tính khả thi cho việc tổ chức thực hiện nghiên cứu trong thực tế. Có thể từ một câu hỏi trọng tâm có vài câu hỏi thứ cấp để giới hạn phạm vi cho từng nghiên cứu cụ thể.

Các “vấn đề” hay câu hỏi nghiên cứu trong chăn nuôi để hình thành nên các đề tài nghiên cứu khoa học thường xuất phát từ các nguồn sau đây:

- Nhà khoa học

Người làm khoa học có cách nghĩ theo kiểu “khoa học” và thường đặt ra những câu hỏi để nghiên cứu. Chẳng hạn, qua quá trình nghiên cứu tài liệu nhà khoa học phát hiện ra các “vấn đề” thắc mắc, các lỗ hổng tri thức. Đôi khi người nghiên cứu thấy một điều gì đó chưa rõ trong những nghiên cứu trước của những người khác và muốn đặt câu hỏi để kiểm chứng lại hay làm sáng tỏ thêm. Cũng có thể trong các hội nghị khoa học có những tranh luận, thậm chí bất đồng, giúp cho các nhà khoa học nhận thấy được những mặt yếu, mặt hạn chế của các công trình mà mình đã và đang nghiên cứu, từ đó nhận định, phân tích lại và chọn lọc để rút ra những “vấn đề” cần nghiên cứu tiếp. Các câu hỏi nghiên cứu cũng có khi chợt xuất hiện trong suy nghĩ của các nhà khoa học. Đó cũng có thể là do tính tò mò hay suy luận chủ quan của nhà khoa học về một điều gì đó mà thấy cần phải nghiên cứu để trả lời.

- Người tiêu dùng và xã hội

Nhu cầu đời sống của người dân không ngừng tăng cao và đa dạng, người tiêu dùng đòi hỏi ngành chăn nuôi ngày càng có thêm những sản phẩm tốt hơn. Các nhà khoa học ngành chăn nuôi cần phải nắm được nhu cầu của người tiêu dùng để hình thành nên những ý tưởng nghiên cứu nhằm đáp ứng được nhu cầu của họ. Ngành chăn nuôi cho dù hiểu theo nghĩa sản xuất (animal production) hay một lĩnh vực khoa học (animal science) thì đều mang tính ứng dụng, do vậy nghiên cứu chăn nuôi trước tiên phải đáp ứng nhu cầu của người tiêu dùng về sản phẩm chăn nuôi. Mặt khác, chăn nuôi cũng có mặt trái của nó và xã hội có thể tạo áp lực buộc các nhà khoa học chăn nuôi phải nghiên cứu để hạn chế những tiêu cực do chăn nuôi gây ra như “vấn đề” chất thải chăn nuôi gây ô nhiễm môi trường hay gia tăng biến đổi khí hậu. Đáp ứng được nhu cầu của người tiêu dùng và xã hội thì nghiên cứu chăn nuôi mới đáp ứng được đòi hỏi hàng đầu đối với một ngành khoa học ứng dụng là “khoa học vị nhân sinh” bền vững.

- Thực tiễn sản xuất chăn nuôi

Vấn đề nghiên cứu cũng có thể được hình thành qua những nhu cầu hay bức xúc nảy sinh trong thực tiễn sản xuất chăn nuôi. Các vấn đề nghiên cứu phải xuất phát trước hết từ nhu cầu thực tiễn, nhằm giải quyết những khó khăn và/hay phát huy lợi thế của sản xuất chăn nuôi trong nước và của từng địa phương cụ thể. Các nhà nghiên cứu chăn nuôi của Việt Nam trước tiên phải tập trung nghiên cứu để giải quyết những “vấn đề” của ngành chăn nuôi Việt Nam. Sản xuất chăn nuôi là nơi ứng dụng kết quả và kiểm chứng giá trị đích thực của một nghiên cứu chăn nuôi.

- Nhà quản lý

Trong quá trình quản lý ngành hay quản lý xã hội, các nhà quản lý có thể thấy những vấn đề trong ngành chăn nuôi cần phải nghiên cứu giải quyết. Tuy nhiên, chỉ có những nhà quản lý thực sự là những “chiến lược gia” có tầm nhìn tổng thể và kinh nghiệm thực tiễn thì mới biết “đặt hàng” cho các nhà khoa học nghiên cứu giải quyết những vấn đề nổi cộm cụ thể cho ngành và cho từng địa phương.

- Đồng nghiệp và tổ chức tài trợ

Các đồng nghiệp, kể cả đồng nghiệp nước ngoài, thường trao đổi ý tưởng về những “vấn đề” cần cùng nhau nghiên cứu giải quyết. Nhiều tổ chức quốc tế cũng thường kêu gọi các đề xuất nghiên cứu liên quan đến chăn nuôi như nghiên cứu bảo tồn nguồn gen vật nuôi, giảm phát thải khí nhà kính từ gia súc Các nghiên cứu từ những ý tưởng như vậy có vai trò rất quan trọng trong việc tăng cường và nâng cao năng lực nghiên cứu cho các nhà khoa học cũng như các cơ quan nghiên cứu chăn nuôi của nước ta, cũng như đóng góp phần bổ sung tri thức chung cho nhân loại.

1.2.2. Cơ sở lý luận và thực tiễn (Tổng quan tài liệu)

Thu thập và nghiên cứu tài liệu là một công việc quan trọng cần thiết cho bất kỳ hoạt động nghiên cứu khoa học nào. Các nhà nghiên cứu khoa học luôn đọc và tra cứu

tài liệu có trước để làm nền tảng cho nghiên cứu khoa học của mình. Đây là nguồn kiến thức quý giá được tích lũy qua quá trình nghiên cứu mang tính lịch sử lâu dài. Mục đích của việc thu thập và nghiên cứu tài liệu nhằm:

- Giúp cho người nghiên cứu có thêm kiến thức rộng và sâu về lĩnh vực đang nghiên cứu, đặc biệt qua nghiên cứu tài liệu xung quanh vấn đề quan tâm nghiên cứu, nhà khoa học có thể biết được *những gì đã và chưa được nghiên cứu* xung quanh vấn đề mình đang quan tâm.

- Nắm được phương pháp của các nghiên cứu đã thực hiện trước đây để có phương pháp luận hay luận cứ chặt chẽ cho nghiên cứu của mình.

- Tránh trùng lặp với các nghiên cứu trước đây, vì vậy đỡ mất thời gian, công sức và tài chính.

- Giúp người nghiên cứu xây dựng luận cứ (bằng chứng) để chứng minh giả thuyết nghiên cứu khoa học.

Khi đã xác định một vấn đề hay câu hỏi nghiên cứu, nhà khoa học cần nghiên cứu tổng quan các tài liệu liên quan đến vấn đề đó để tìm logic cho công trình nghiên cứu, trả lời câu hỏi “*tại sao lại làm nghiên cứu này?*” để cho thấy “tính cấp thiết” của công trình nghiên cứu vì nó có ý nghĩa gì về khoa học và thực tiễn.

Nghiên cứu tổng quan tài liệu có thể tiến hành sau khi đã có câu hỏi nghiên cứu sơ bộ. Tuy nhiên, qua nghiên cứu tài liệu nhà khoa học cũng có thể thấy câu hỏi nghiên cứu mình đặt ra đã có đáp án rồi và nó không còn là “câu hỏi nghiên cứu” nữa. Việc nghiên cứu tổng quan tài liệu khi đó chỉ có ý nghĩa làm gia tăng kiến thức cho nhà khoa học về vấn đề mình đang quan tâm. Ngược lại, nếu tổng quan tài liệu cho thấy câu hỏi đặt ra thực sự chưa có câu trả lời từ các công trình nghiên cứu khoa học trước đây thì đó chính là điểm xuất phát cho một công trình nghiên cứu khoa học mới có ý nghĩa vì nó sẽ đem lại kiến thức mới, giúp ta hiểu được những điều muốn hiểu.

Mặt khác, trên cơ sở tổng quan những kết quả nghiên cứu đã được công bố, những vấn đề nóng đang đặt ra trong thực tiễn chăn nuôi, những đề tài nghiên cứu của chính mình liên quan trực tiếp đến vấn đề đang được quan tâm nghiên cứu, từ việc đánh giá có biện luận những điểm mạnh và điểm yếu của các nghiên cứu trước, cần tìm ra được đâu là lỗ hổng/khoảng trống kiến thức hiện tại (knowledge gap) về vấn đề nghiên cứu để dẫn dắt cách giải quyết vấn đề đó trong nghiên cứu của mình thông qua một hoặc một số câu hỏi nghiên cứu cụ thể nhằm vào lỗ hổng kiến thức đó hay đáp ứng những vấn đề bất cập do thực tiễn sản xuất đặt ra (mà chưa có câu trả lời hay giải pháp hữu hiệu). Nghiên cứu nhằm trả lời câu hỏi đó (có thể có nhiều hơn 1 câu hỏi) sẽ cho ra kết quả mới, giúp góp phần giải quyết vấn đề đang quan tâm.

Cách viết Cơ sở lý luận và thực tiễn hay Tổng quan tài liệu sẽ được trình bày cụ thể trong các mục đề cập đến *Đề cương nghiên cứu, Bài báo khoa học và Luận án*.

1.2.3. Giả thuyết khoa học

Giả thuyết là câu trả lời ướm thử hoặc là sự tiên đoán để trả lời cho câu hỏi hay “vấn đề” nghiên cứu. Sau khi xác định được “vấn đề” hay câu hỏi nghiên cứu có ý nghĩa, người nghiên cứu hình thành ý tưởng nghiên cứu để tìm câu trả lời cho câu hỏi nghiên cứu hoặc sự giải thích cho “vấn đề” chưa biết. Câu trả lời hay giải thích này còn gọi là sự tiên đoán khoa học hay *giả thuyết khoa học*. Người nghiên cứu hình thành một cơ sở lý luận để xây dựng giả thuyết khoa học trên cơ sở những quan sát bước đầu, những tình huống đặt ra (câu hỏi những cơ sở lý thuyết (tham khảo tài liệu, kiến thức đã có,...), sự tiên đoán và những dự kiến tiến hành thực nghiệm. *Giả thuyết khoa học là một phát biểu mang tính tiên lượng về mối quan hệ giữa hai hay nhiều biến*. Điều đó có thể đúng nhưng cũng có thể sai, nên cần làm thí nghiệm để kiểm chứng. Giả thuyết khoa học chính là sự dự đoán chủ quan về kết cục của một nghiên cứu. Nó giúp cho người nghiên cứu có động cơ, hướng đi hay cách tiếp cận để thiết kế phương pháp nghiên cứu. Ví dụ, giả thuyết cho rằng *một tỷ lệ thức ăn tinh cao trong khẩu phần của bò sữa dẫn đến tỷ lệ thụ thai thấp*. Để biết được điều này đúng hay sai thì cần làm thí nghiệm để kiểm chứng.

Mỗi giả thuyết khoa học chứa 2 loại biến (các quan sát có thể thể hiện được bằng các giá trị khác nhau):

- *Biến độc lập*: là biến mà người nghiên cứu (hay tự nhiên) có thể kiểm soát/điều chỉnh được;

- *Biến phụ thuộc*: Là biến cần đo đạc xác định hay quan sát trên đối tượng thí nghiệm, được cho là chịu ảnh hưởng của biến độc lập.

Các cấu trúc đặc trưng của một giả thuyết có thể thuộc một trong các dạng sau đây:

- *Điều kiện*: Nếu điều kiện X (biến độc lập) được thực hiện thì kết cục Y (biến phụ thuộc) sẽ xảy ra.

- *Tương quan*: Giá trị quan sát được của biến số B (biến phụ thuộc) có liên quan đến những thay đổi về giá trị của biến số A (biến độc lập).

- *Nhân - Quả*: Giá trị của biến số A (biến độc lập) quyết định giá trị của biến số B (biến phụ thuộc).

Điều quan trọng trong cách đặt giả thuyết là phải đảm bảo có thể thực hiện thí nghiệm để kiểm chứng giả thuyết “đúng” hay “sai” của giả thuyết. Vì vậy, để xây dựng một giả thuyết cần trả lời được các câu hỏi sau:

- (1) Giả thuyết có thể tiến hành thực nghiệm để kiểm chứng được không?

- (2) Các biến hay các yếu tố nào cần được nghiên cứu?

- (3) Phương pháp thí nghiệm nào (trong phòng thí nghiệm hay điều tra thực địa,...) cần được sử dụng trong nghiên cứu?

- (4) Các chỉ tiêu nào cần theo dõi trong thí nghiệm?

- (5) Phương pháp xử lý số liệu nào cần dùng để bác bỏ hay chấp nhận giả thuyết?

(6) Do vậy, một giả thuyết khoa học hợp lý cần có các đặc điểm chính sau đây:

- Dựa trên quan sát hiện tại;
- Chưa được kiểm chứng khẳng định;
- Có khả năng đúng hay sai (nếu giả thuyết đặt ra đã được khẳng định là chỉ có đúng thì không cần nghiên cứu thêm);
- Có thể kiểm chứng bằng thí nghiệm (vì nếu không thì không thể tổ chức nghiên cứu để kiểm chứng).

Tóm lại, giả thuyết là một *câu trả lời ước thử* cho một câu hỏi nghiên cứu trên cơ sở suy luận của nhà khoa học, nó *có thể đúng hay sai*. Chính vì thế giả thuyết là cơ sở để hình thành một công trình nghiên cứu khoa học để kiểm chứng và có thể hình thành nên lý thuyết khoa học mới.

1.2.4. Thí nghiệm chăn nuôi

Để kiểm chứng một giả thuyết khoa học trong chăn nuôi thông thường cần phải làm thí nghiệm. Muốn có được thí nghiệm khả thi và có giá trị trước hết phải thiết kế thí nghiệm đúng, trong đó trả lời được các câu hỏi: *làm gì? làm như thế nào/thu thập và phân tích số liệu bằng cách nào?* để có thể ủng hộ hay bác bỏ giả thuyết nghiên cứu. Sau đây là một số nội dung cơ bản cần quan tâm khi thiết kế và tiến hành làm thí nghiệm nghiên cứu chăn nuôi.

a. Mô hình thiết kế thí nghiệm

Lựa chọn mô hình *thiết kế thí nghiệm* và dự kiến mô hình *phân tích thống kê* số liệu thí nghiệm là một phần của quá trình lựa chọn phương pháp nghiên cứu. Biết được cách phân tích số liệu nào sẽ được áp dụng và cách thức diễn giải kết quả thì sẽ có cách thiết kế phương pháp thu thập số liệu đúng cũng như số lượng quan sát cần phải có. Khi thiết kế thí nghiệm, người nghiên cứu nên tạo ra một bộ số liệu mô phỏng (dummy data), tính toán thử trên máy tính để kiểm tra mô hình thiết kế thí nghiệm và phương pháp xử lý thống kê có thực hiện được hay không.

Nội dung chi tiết của thiết kế/kế hoạch thí nghiệm phải được thể hiện trong *Đề cương nghiên cứu/Thuyết minh đề tài* (xem mục 1.3). Mục này không đi sâu về phương pháp lựa chọn mô hình thiết kế thí nghiệm chăn nuôi vì đã đề cập trong Giáo trình Thiết kế thí nghiệm (Nguyễn Đình Hiền và Đỗ Đức Lực, 2007), nghiên cứu sinh có thể tham khảo lại Giáo trình đó. Một số nội dung quan trọng liên quan đến thiết kế và tiến hành thí nghiệm chăn nuôi được trình bày trong các tiểu mục sau đây.

b. Đối tượng nghiên cứu và mẫu

Công việc đầu tiên phải làm để tiến hành một thí nghiệm là xác định *đối tượng nghiên cứu*. Đó là quần thể đích (target population) bao gồm nhiều cá thể mang các đặc điểm tương tự nhau. Trong nghiên cứu, có nhiều lý do không thể quan sát hay làm thí nghiệm được trên toàn bộ các cá thể trong quần thể đích, mà chỉ chọn một số lượng đủ

các cá thể đại diện hay còn gọi là mẫu. Mẫu (sample) là một phần hoặc một tập hợp nhỏ các cá thể của quần thể đích được chọn làm đại diện để tiến hành nghiên cứu. Sau đó các dữ liệu thu được từ mẫu sẽ được phân tích để đưa ra các kết luận khái quát chung cho đối tượng nghiên cứu. Các kết luận này được gọi là “kết luận thống kê”. Để các kết luận đưa ra đúng cho đối tượng nghiên cứu thì mẫu phải “phản ánh” được quần thể (hay còn nói là mẫu phải “đại diện”/“điển hình” cho quần thể), không được thiên về phía “tốt” hay thiên về phía “xấu”.

Do vậy, khi đã xác định được đối tượng nghiên cứu, để bố trí thí nghiệm thì bước tiếp theo là thiết lập khung mẫu. Thiết lập khung mẫu là xác định các cá thể đại diện cho đối tượng nghiên cứu làm mẫu nghiên cứu, xác định cỡ mẫu (dung lượng mẫu) và phương pháp lấy mẫu. Nếu thiết lập khung mẫu sai thì mẫu chọn sẽ không mang tính đại diện và kết quả thu được sẽ không phản ánh đúng đối tượng nghiên cứu.

- Xác định các cá thể làm mẫu thí nghiệm

Các cá thể được chọn làm mẫu thí nghiệm phải mang được những đặc tính đại diện cho đối tượng nghiên cứu. Những đặc tính không phải là yếu tố thí nghiệm phải thể hiện tương đối “đồng đều” giữa các cá thể khi chọn làm thí nghiệm. Trong trường hợp không thể chọn được các cá thể “đồng đều” về các yếu tố phi thí nghiệm thì phải nhận dạng các những yếu tố sai khác nào có thể gây “nhiều” kết quả để đo đạc được và tìm giải pháp thống kê loại trừ ảnh hưởng “nhiều”. Một số đặc tính (chủ yếu là các chỉ tiêu theo dõi thí nghiệm) có thể thay đổi trong quá trình thí nghiệm dưới ảnh hưởng của các yếu tố thí nghiệm và sẽ được phản ánh trong “kết quả” thí nghiệm.

- Xác định cỡ mẫu

Mục đích của việc xác định cỡ mẫu (dung lượng mẫu) là giảm thiểu công lao động và chi phí làm thí nghiệm mà không làm mất đi các đặc tính của mẫu cũng như độ tin cậy của số liệu đại diện cho quần thể và kết quả nghiên cứu. Do vậy, cần tính toán để xác định một kích cỡ mẫu tối thiểu mà vẫn đánh giá được tương đối chính xác quần thể. Chọn cỡ mẫu quá lớn hoặc lớn hơn mức tối thiểu thì tốn kém, còn chọn cỡ mẫu dưới mức tối thiểu sẽ dẫn đến kết luận thiếu chính xác.

Đối với thí nghiệm chăn nuôi, số động vật thí nghiệm phải đủ sao cho các đặc tính riêng biệt của từng cá thể không làm ảnh hưởng đến kết quả. Nếu số động vật trong thí nghiệm quá ít thì độ tin cậy của kết quả thu được từ thí nghiệm sẽ không cao. Ngược lại, nếu số động vật quá nhiều thì có thể gây lãng phí. Để đạt được độ chính xác không phải lúc nào cũng cần số lượng động vật thí nghiệm quá lớn. Nếu quá nhiều động vật tham gia thí nghiệm thì có thể gây ra nhiều khó khăn trong quá trình theo dõi từng cá thể, khó khăn khi chúng ta muốn tạo ra các điều kiện đồng nhất của thí nghiệm cho mọi cá thể, chẳng hạn như nơi nhốt, thời gian cho động vật ăn,... Điều đó sẽ làm giảm độ chính xác về mặt kỹ thuật của thí nghiệm.

Dung lượng mẫu cần thiết còn phụ thuộc vào chất lượng của động vật tham gia thí nghiệm. Động vật tham gia thí nghiệm có độ đồng đều cao thì giảm số lượng xuống và ngược lại. Độ tuổi của vật nuôi cũng đóng vai trò quan trọng trong quá trình chọn dung

lượng mẫu. Động vật càng non thì số lượng cần phải tăng lên và ngược lại, bởi vì đối với loại động vật này mức độ biến động rất lớn (cả về mặt sinh lý và ngoại hình). Ngoài ra, dung lượng mẫu còn phụ thuộc vào từng loại vật nuôi. Mỗi loại vật nuôi có những đặc điểm riêng, vì vậy trong quá trình thiết kế thí nghiệm cũng phải chú ý đến yếu tố này. Cuối cùng, kết quả mong đợi của thí nghiệm (sự chênh lệch giữa các công thức thí nghiệm) cũng ảnh hưởng rất nhiều đến dung lượng mẫu.

Một số cách tiếp cận để tính dung lượng mẫu như sau:

* Cách tiếp cận thứ nhất

- Số nghiệm thức (a)

- Mức độ đồng đều của tính trạng cần nghiên cứu (σ^2). Theo qui luật, nếu như cỡ mẫu $n < 30$, chúng ta có thể tính σ từ độ lệch chuẩn mẫu S theo công thức. Ngoài ra cũng có thể tính σ từ những quần thể tương tự hoặc từ các thử nghiệm thí điểm, hoặc thậm chí phải phỏng đoán.

- Sai lầm loại I (α)

- Sai lầm loại II (β) hoặc độ mạnh (lực lượng) của phép thử ($1 - \beta$)

- Chênh lệch bé nhất giữa 2 giá trị trung bình bất kỳ để phát hiện sai khác (d) (nếu có).

Các trường hợp tính toán cụ thể theo cách tiếp cận này như sau:

+ Trường hợp ước tính một giá trị trung bình

Trước khi xác định cỡ mẫu, phải mặc nhận mẫu được chọn từ một quần thể có sự phân phối chuẩn về các chỉ tiêu nghiên cứu. Để xác định cỡ mẫu tối thiểu cần phải đánh giá trung bình quần thể μ . Khi thu thập số liệu từ mẫu thì tính được trung bình mẫu. Trung bình mẫu này khác với trung bình quần thể μ . Sự khác nhau giữa mẫu và quần thể được xem là sai số (d). Dung lượng mẫu cần thiết để giá trị trung bình cộng ước tính khác μ không quá d khi có phân phối chuẩn $N(\mu, \sigma^2)$ và mức tin cậy $P = 1 - \alpha$ được tính dựa vào công thức sau:

$$n \geq \frac{(z_{1-\alpha/2})^2 \times \sigma^2}{d^2}$$

Ví dụ: Cần quan sát bao nhiêu bò sữa để ước tính được năng suất trong chu kỳ tiết sữa 305 ngày với mức độ tin cậy 95% nằm trong khoảng $\pm 75\text{kg}$ so với giá trị thực của quần thể. Biết rằng sản lượng sữa có phân phối chuẩn với $\sigma = 500\text{kg}$.

$$\text{Cần thiết: } n \geq \frac{(z_{1-\alpha/2})^2 \times \sigma^2}{d^2} = \frac{1,96^2 \times 500^2}{75^2} = 170,74$$

Như vậy cần ít nhất 171 bò sữa để thỏa mãn điều kiện bài toán.

+ Trường hợp ước tính một tỷ lệ

Tính dung lượng mẫu (n) cần thiết để tỷ lệ ước tính $\hat{p}$ khác không quá d so với tỷ lệ thực π . Nếu biết tỷ lệ hiện hành p (prevalance) và kiểm định ở mức tin cậy $P = 1 - \alpha$, tính n dựa vào công thức sau:

$$n \geq \frac{(z_{1-\alpha/2})^2 \times p(1-p)}{d^2}$$

Lưu ý: Tỷ lệ hiện hành p có thể tìm được thông qua các tài liệu, các nghiên cứu trước hoặc xuất phát từ kinh nghiệm và sự hiểu biết của người nghiên cứu. Nếu khi tiến hành thí nghiệm không có thông tin về tỷ lệ lưu hành, ta sẽ chọn $p = 0,5$. Khi đó n được tính theo công thức:

$$n \geq \frac{(z_{1-\alpha/2})^2}{4d^2}$$

Ví dụ: Cần dung lượng mẫu bao nhiêu để xác định tỷ lệ hiện nhiễm một loại vi khuẩn trên thân thịt lợn ở một lò mổ với ước tính chênh lệch không quá 5%. Biết rằng tỷ lệ hiện hành $p = 0,2$ và kiểm định ở mức tin cậy 95%.

$$\text{Cần thiết: } n \geq \frac{(z_{1-\alpha/2})^2 \times p(1-p)}{d^2} = \frac{1,96^2 \times (1-0,2)}{0,05^2} = 245,86$$

Như vậy cần khảo sát ít nhất 246 thân thịt.

+ Trường hợp so sánh 2 giá trị trung bình

Tính dung lượng mẫu (n) cần thiết (đối với mỗi nghiệm thức) để phát hiện được sự sai khác nếu chênh lệch giữa 2 giá trị trung bình là d, sai lầm loại I và loại II ở mức tương ứng là α và β . Giả sử số liệu có phân phối chuẩn. Phương sai của tính trạng nghiên cứu là σ^2 . Dung lượng mẫu được tính theo công thức:

$$n \geq \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2} 2\sigma^2$$

Ví dụ: Muốn thiết kế một thí nghiệm để so sánh sản lượng sữa của dê Bách Thảo ở 2 công thức thí nghiệm với yêu cầu $\alpha = 0,05$; $\beta = 0,2$; chênh lệch mong đợi 30kg sữa, biết $\sigma = 50\text{kg}$.

$$\text{Cần thiết: } n \geq \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2} 2\sigma^2 = \frac{(1,96 + 0,84)^2}{30^2} \approx 44$$

Như vậy cần ít nhất 44 con dê cho mỗi công thức thí nghiệm.

+ Trường hợp so sánh hai tỷ lệ

Tính dung lượng mẫu (n) cần thiết (đối với mỗi nghiệm thức) để phát hiện được sự sai khác giữa 2 tỷ lệ, sai lầm loại I và loại II ở mức tương ứng là α và β . Tỷ lệ ở quần thể 1 và 2 tương ứng là p_1 và p_2 . Dung lượng mẫu được tính theo công thức:

$$n \geq \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_2)^2} (p_1 q_1 + p_2 q_2)$$

+ Trường hợp so sánh nhiều giá trị trung bình

Nếu ảnh hưởng của nghiệm thức ít, muốn phát hiện được sự ảnh hưởng này đòi hỏi dung lượng mẫu phải lớn. Bên cạnh đó, các giá trị sai lầm loại I và độ mạnh của phép thử tương ứng là α và $1 - \beta$ cũng ảnh hưởng rất nhiều đến dung lượng mẫu cần thiết.

Dung lượng mẫu cần thiết đối với mỗi nghiệm thức (n) để phát hiện sự sai khác (nếu có) khi chênh lệch bé nhất giữa 2 giá trị trung bình bất kỳ là d, số nghiệm thức là a và phương sai của tính trạng nghiên cứu là σ^2 được tính dựa trên công thức dưới đây:

$$\phi^2 = \frac{nd^2}{2a\sigma^2}$$

Tham số ϕ^2 được đề cập chi tiết ở Chương 4 của Giáo trình Thiết kế thí nghiệm (Nguyễn Đình Hiền và Đỗ Đức Lực, 2007). Ở đây ta sẽ sử dụng đường cong cho sẵn ở phần phụ lục để tìm dung lượng mẫu cần thiết.

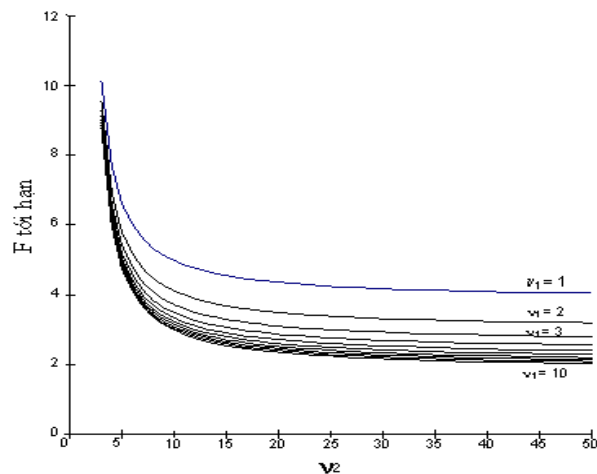
Ví dụ: Nghiên cứu tăng khối lượng (g/ngày) của lợn nuôi vỗ béo đến 5 tháng tuổi ở 3 công thức thí nghiệm. Hãy xác định dung lượng mẫu (n) cần thiết để phát hiện sự sai khác giữa các nghiệm thức nếu có. Biết rằng sự chênh lệch giữa 2 giá trị trung bình lúc kết thúc thí nghiệm là 40g, tăng khối lượng có phân phối chuẩn với phương sai $\sigma^2 = 480$. Sử dụng công thức nêu trên cùng với các đường cong ở phần phụ lục ta có thể tìm ra dung lượng mẫu cần thiết ở các mức chính xác tương ứng:

Nếu $\alpha = 0,05$; $1 - \beta = 0,80 \Rightarrow n = 7$; Nếu $\alpha = 0,05$; $1 - \beta = 0,90 \Rightarrow n = 9$

Nếu $\alpha = 0,01$; $1 - \beta = 0,80 \Rightarrow n = 10$; Nếu $\alpha = 0,01$; $1 - \beta = 0,90 \Rightarrow n = 12$

* Cách tiếp cận thứ hai để xác định được dung lượng mẫu cần thiết là dựa vào số bậc tự do tối thiểu của sai số ngẫu nhiên.

Giả sử so sánh hai hoặc nhiều nghiệm thức với nhau ta mong muốn bậc tự do của sai số ngẫu nhiên ≥ 20 bởi vì với bậc tự do của sai số ngẫu nhiên bé thì giá trị tới hạn của F rất lớn, nhưng nó sẽ giảm rất nhanh khi bậc tự do này tăng lên. Khi bậc tự do sai số ngẫu nhiên lớn hơn 20 thì giá trị F giảm rất ít. Trong đồ thị dưới đây, v_1 và v_2 tương ứng với bậc tự do của nghiệm thức và bậc tự do của sai số ngẫu nhiên sẽ minh họa điều này.



Đồ thị 1.1. Giá trị tối hạn của phân phối F với bậc tự do v_1 , v_2 và $\alpha = 0,05$

Sử dụng quy tắc bậc tự do tối thiểu trên ta có thể tính dung lượng mẫu cần thiết cho ví dụ sau:

Ví dụ: Thiết kế thí nghiệm kiểu khối ngẫu nhiên hoàn toàn (CRD) với 1 lần lặp lại có số nghiệm thức $a = 5$. Cần bao nhiêu khối (b) và bao nhiêu cá thể làm đơn vị thí nghiệm (N)?

Ta cần số bậc tự do của sai số ngẫu nhiên là $df = (b - 1) \times (a - 1) \geq 20$

Như vậy, ta cần có $(b - 1) \times (5 - 1) \geq 20$ hay số khối là $b \geq 6$.

Vì vậy, tổng số cá thể cần tối thiểu là $N = a \times b = 5 \times 6 = 30$

Ngoài các cách tiếp cận nêu trên, các nhà nghiên cứu cũng đưa ra các khuyến cáo khác nhau để dựa vào nó mà có thể tìm ra dung lượng mẫu phù hợp. Ví dụ: Trong nghiên cứu về đại gia súc, Preston (1995) cho rằng số động vật trong một công thức thí nghiệm không được ít hơn 3 và bậc tự do của sai số ngẫu nhiên ít nhất là 15. Trong các nghiên cứu về đại gia súc và lợn, Ovesiannhicov (1976) khuyến cáo số động vật trong một nghiệm thức ít nhất là 6 và thuận lợi hơn nếu con số này là 12.

- Phương pháp lấy mẫu

Mục đích của tất cả các phương pháp lấy mẫu là có được mẫu đại diện cho đối tượng nghiên cứu. Phương pháp chọn mẫu thí nghiệm rất quan trọng có liên quan tới sự biến động hay độ đồng đều của mẫu. Có hai phương pháp chọn mẫu: (1) Chọn mẫu không ngẫu nhiên và (2) Chọn mẫu ngẫu nhiên.

- Chọn mẫu không ngẫu nhiên

Phương pháp chọn mẫu không ngẫu nhiên là cách lấy mẫu trong đó các cá thể của mẫu được chọn không ngẫu nhiên hay các cá thể của quần thể mục tiêu không có xác suất lựa chọn giống nhau. Điều này thể hiện trong cách chọn mẫu như sau:

+ Các đơn vị mẫu được tự lựa chọn không theo một phương pháp cụ thể nào.

+ Các đơn vị mẫu dễ đạt được hoặc dễ tiếp cận. Ví dụ, chọn những con gặp trên đường đi.

+ Các đơn vị mẫu được chọn theo lý do kinh tế. Ví dụ, chọn những mẫu sẵn có không mất tiền mua.

+ Các đơn vị mẫu được quan tâm bởi người nghiên cứu trong cách “điền hình” của quần thể mục tiêu. Ví dụ, người nghiên cứu chỉ quan tâm đến các cá thể điền hình trong quần thể nghiên cứu, để so sánh với các cá thể khác.

+ Các đơn vị mẫu được chọn mà không có sự thiết kế rõ ràng. Ví dụ, chọn 50 con vật đầu tiên gặp được.

Phương pháp chọn mẫu không ngẫu nhiên như trên thường có độ tin cậy thấp. Mức độ chính xác của cách chọn mẫu không xác suất tùy thuộc vào sự phán đoán, cách nhìn, kinh nghiệm của người nghiên cứu, sự may mắn hoặc dễ dàng và không có cơ sở thống kê trong việc chọn mẫu.

- Chọn mẫu ngẫu nhiên

Việc chọn mẫu ngẫu nhiên hay xác suất là cách lấy chọn các cá thể của mẫu mà các cá thể thuộc đối tượng nghiên cứu có cơ hội lựa chọn như nhau. Nếu như có một số cá thể có cơ hội xuất hiện nhiều hơn thì sự lựa chọn không phải là ngẫu nhiên. Để tối ưu hóa mức độ chính xác, người nghiên cứu thường sử dụng phương pháp lấy mẫu ngẫu nhiên. Có các phương pháp chọn mẫu ngẫu nhiên sau:

+ Chọn mẫu ngẫu nhiên đơn giản

Cách đơn giản nhất của việc chọn các cá thể của mẫu trong cách chọn mẫu ngẫu nhiên là sử dụng xác suất. Việc lựa chọn n các cá thể từ một quần thể sao cho các cá thể có cơ hội bằng nhau hay một xác suất bằng nhau trong phương pháp này.

Ví dụ: Một trại bò có 1.000 con bò vắt sữa, người nghiên cứu muốn chọn ra 100 bò vắt sữa để nghiên cứu về tình trạng sức khỏe của đàn bò này. Theo cách chọn mẫu đơn giản thì chỉ cần viết số hiệu của 1.000 con bò đó vào trong mẫu giấy nhỏ, sau đó bỏ tất cả vào trong một cái thùng và rồi rút ngẫu nhiên ra 100 mẫu giấy. Như vậy, mỗi con có một cơ hội lựa chọn như nhau và xác suất chọn ngẫu nhiên một con trên dễ dàng được tính. Trong ví dụ trên ta có quần thể $N = 1.000$ và cỡ mẫu $n = 100$. Như vậy, mỗi con được chọn trong cách lấy mẫu ngẫu nhiên sẽ có xác suất là $n \times 100/N$ (%) hay $100 \times 100/1000 = 10\%$.

Một cách chọn mẫu ngẫu nhiên khác là sử dụng bảng số ngẫu nhiên trong sách thống kê thí nghiệm hoặc cách chọn số ngẫu nhiên bằng các chương trình thống kê trên máy tính.

+ Chọn mẫu phân lớp

Chọn mẫu phân lớp được thực hiện khi quần thể mục tiêu được chia thành các nhóm hay phân lớp. Trong phương pháp lấy mẫu phân lớp, tổng quần thể (N) đầu tiên được chia ra thành L lớp (các quần thể phụ) là $N_1, N_2 \dots N_L$.

Để áp dụng kỹ thuật chọn mẫu phân lớp thì trước tiên người nghiên cứu cần nắm các thông tin và các số liệu nghiên cứu trước đây có liên quan đến cách lấy mẫu phân lớp. Sau đó, người nghiên cứu sẽ xác định cỡ mẫu và chọn ngẫu nhiên các cá thể trong mỗi lớp.

+ Chọn mẫu hệ thống

Trong chọn mẫu hệ thống, cỡ mẫu n được chọn (có phương pháp tính xác suất tương tự) từ một quần thể N . Cách lấy mẫu hệ thống là khung mẫu giống như là 1 “hàng” của các đơn vị mẫu, và mẫu như là một chuỗi liên tiếp của các điểm số có khoảng cách bằng nhau theo hàng dọc.

Ví dụ: Chọn 1 cá thể trong mỗi nhóm có 10 cá thể của một quần thể có 10 nhóm (tổng cá thể của quần thể là 100, đánh số cá thể từ 1 đến 100). Lúc này nhóm 1 được đánh số từ 1 đến 10; nhóm 2 từ 11 đến 20; nhóm 3 từ 21 đến 30; ...nhóm 10 từ 91 đến 100. Trước tiên cần sắp xếp thứ tự các đơn vị mẫu (ví dụ theo thứ tự gia tăng trong trường hợp này). Sau đó chọn điểm đầu tiên bất kỳ có giá trị <10 (ví dụ chọn ngẫu nhiên một số trong khoảng từ 1 đến 10 là 7. Số cá thể tiếp theo sẽ cộng thêm là 10. Như vậy các cá thể được chọn sẽ có số thứ tự là 7, 17, 27, 37, ... 97.

+ Chọn mẫu chỉ tiêu

Trong cách chọn mẫu chỉ tiêu, quần thể nghiên cứu được phân nhóm hoặc phân lớp như cách chọn mẫu phân lớp. Các đối tượng nghiên cứu trong mỗi nhóm được lấy mẫu theo tỷ lệ đã biết và sau đó tiến hành phương pháp chọn mẫu không xác suất. Để thiết lập mẫu chỉ tiêu thì người nghiên cứu cần biết các số liệu, thông tin trong quần thể mục tiêu để phân chia các chỉ tiêu muốn kiểm soát.

Ưu điểm của lấy mẫu chỉ tiêu áp dụng trong một vài nghiên cứu là dễ thực hiện và chi phí nghiên cứu tương đối thấp (do không cần phải thiết lập khung mẫu). Nhược điểm là không đại diện cho toàn bộ quần thể, do lấy mẫu không xác suất và vì vậy mức độ tin cậy phụ thuộc vào kinh nghiệm hay sự phán đoán của người nghiên cứu. Chính vì thế chọn mẫu chỉ tiêu ít được áp dụng trong các nghiên cứu chăn nuôi, nhưng đôi khi được sử dụng trong một vài nghiên cứu nhỏ mang các đặc tính quan sát.

+ Chọn mẫu theo không gian

Người nghiên cứu có thể sử dụng cách lấy mẫu này khi hiện tượng, sự vật được quan sát có sự phân bố theo không gian (các đối tượng khảo sát trong khung mẫu có vị trí không gian 2 hoặc 3 chiều). Ví dụ: lấy mẫu không khí trong chuồng nuôi. Lấy mẫu theo sự phân bố này yêu cầu có sự giống nhau về không gian qua các phương pháp ngẫu nhiên, hệ thống và phân lớp. Kết quả chọn mẫu có thể được biểu diễn như một loạt các điểm trong không gian hai chiều, giống như là bản đồ.

c. Mức yếu tố và nghiệm thức

Các phần tử riêng biệt khác nhau trong cùng một yếu tố thí nghiệm được gọi là mức. Ví dụ, ta có một yếu tố thí nghiệm là kiểu gen Halothane ở lợn thì sẽ có 3 phần tử khác nhau tương ứng với 3 kiểu gen (NN, Nn, nn) hay còn được gọi là 3 mức. Hoặc khi nghiên cứu ảnh hưởng của protein đến sản lượng sữa bò ta có thể nghiên cứu ở 3 mức protein khác nhau (ví dụ 14, 16 và 18%).

Đối với các thí nghiệm có biến độc lập (yếu tố thí nghiệm) vốn có bản chất là biến liên tục nhưng được chia ra nhiều mức khác nhau thì nên có càng nhiều mức càng tốt. Ví dụ, với 30 con vật thí nghiệm, sẽ thu được nhiều thông tin về phản ứng với một nhân tố (ví dụ, hàm lượng năng lượng trong khẩu phần) nếu bố trí 5 mức với 6 con/mức hơn là 3 mức với 10 con/mức. Bố trí nhiều mức như vậy vẫn đảm bảo được độ chính xác cao đồng thời thu được nhiều thông tin hơn về dạng phản ứng (tuyến tính, parabol hay phi tuyến tính khác). Qua đó ta có thể tìm được mức phản ứng tối ưu của một nhân tố.

Một tổ hợp các mức của các yếu tố thí nghiệm được gọi là một nghiệm thức hay công thức hay lô thí nghiệm. Trong các thí nghiệm, số nghiệm thức có thể bố trí phụ thuộc vào bản chất của thí nghiệm và các nguồn lực có thể huy động. Trong thí nghiệm một nhân tố thì số nghiệm thức bằng chính số mức của nhân tố thí nghiệm đó. Trong thí nghiệm nhiều nhân tố thì số nghiệm thức bằng số tổ hợp các mức của các nhân tố thí nghiệm. Ví dụ, trong nghiên cứu ảnh hưởng của protein ở 3 mức khác nhau đến sản lượng sữa của bò ở trên ta có 3 công thức thí nghiệm tương ứng với 3 mức protein; tuy nhiên, nếu có đủ gia súc thí nghiệm và các nguồn lực khác thì có thể thêm yếu tố thứ 2 là thức ăn tinh trong khẩu phần với 2 mức, lúc này sẽ có tất cả sẽ có $3 \times 2 = 6$ nghiệm thức.

d. Đơn vị thí nghiệm và số lần lặp lại

Một thí nghiệm sử dụng các mẫu của đối tượng nghiên cứu làm đơn vị thí nghiệm. Đơn vị thí nghiệm là *đơn vị thực hiện nhỏ nhất ứng với một công thức thí nghiệm*. Đơn vị thí nghiệm trong chăn nuôi - thú y và thủy sản thường là từng cá thể động vật; tuy nhiên, cũng có lúc lại là một nhóm động vật. Ví dụ, khi nghiên cứu tiêu tốn thức ăn cho một kg tăng khối lượng của gà thịt nhưng trong thực tế không thể theo dõi được lượng thức ăn thu nhận của từng con mà chỉ biết được số thức ăn thu nhận được của từng nhóm gà; tức là từ mỗi nhóm gà như vậy chỉ có một quan sát duy nhất. Đây cũng chính là điều mà các nhà nghiên cứu hay nhầm lẫn. Nếu đơn vị thí nghiệm là một cá thể thì sau khi cân đo ta có được một số liệu hay một quan sát (observation). Nếu đơn vị thí nghiệm là một nhóm gồm nhiều cá thể thì có thể cân đo chung cho cả nhóm hoặc lấy một số cá thể nhất định trong nhóm để cân đo sau đó suy ra một số liệu chung cho đơn vị thí nghiệm. Tương tự như dữ liệu cá thể trong trường hợp trên, số liệu của các nhóm có thể lưu trữ để đánh giá sai số của đơn vị thí nghiệm.

Trong một thí nghiệm, đơn vị thí nghiệm thường được lặp lại một số lần nhất định ở mỗi nghiệm thức, trừ trường hợp đặc biệt. Số lần lặp lại chính là số đơn vị thí nghiệm có trong mỗi nghiệm thức. Thông thường người ta bố trí số lần lặp lại bằng nhau ở các nghiệm thức vì khi số lần lặp của các nghiệm thức bằng nhau thì có thể đưa ra các mô hình phân tích số liệu thuận tiện và đơn giản. Nếu số lần lặp không bằng nhau thì phải sử dụng cách tính phức tạp như mô hình hồi quy nhiều biến tổng quát, kèm theo đó là việc kiểm định các giả thiết, đặc biệt việc tính các kỳ vọng của các trung bình bình phương, sẽ gặp nhiều khó khăn.

Trong thực tế, số lần lặp thường được bố trí lúc đầu thì bằng nhau ở các nghiệm thức, nhưng trong quá trình thí nghiệm có thể không thu thập được đầy đủ dữ liệu vì có

thể có một số động vật bị chết hoặc bị loại thải do không đáp ứng được các yêu cầu. Số lượng động vật thí nghiệm sống đến khi kết thúc thí nghiệm phụ thuộc vào từng loại thí nghiệm và loài vật nuôi khác nhau. Nếu mất ít dữ liệu, có thể tìm cách thay thế dữ liệu bị mất bằng tổ hợp của các dữ liệu còn lại theo một công thức cụ thể, kèm theo sự điều chỉnh của các bậc tự do tương ứng; ngược lại, phải coi như số lần lặp khác nhau và dùng mô hình hồi quy tổng quát để phân tích thống kê. Việc này hiện nay thực hiện không quá khó nhờ có sự trợ giúp của các phần mềm thống kê mạnh như SAS, SPSS, SPLUS, MINITAB,...

Số lần lặp lại càng nhiều càng tốt vì càng dễ phát hiện ra sự sai khác giữa các nghiệm thức nhờ có nhiều độ tự do (degrees of freedom) trong phân tích thống kê. Tuy nhiên, các điều kiện thí nghiệm luôn bị hạn chế, nhất là trong lĩnh vực chăn nuôi, do vậy phải tính đến yếu tố kinh tế khi sử dụng các nguồn lực cho thí nghiệm. Nguyên tắc bao trùm nhất trong thí nghiệm chăn nuôi là không được phép bố trí dưới 3 lần lặp lại (replicates) cho mỗi nghiệm thức.

Có thể ước tính chính xác hơn số lần lặp lại cần thiết để phát hiện được một tỷ lệ sai khác mong đợi với phép thử T theo công thức sau:

$$\text{Tỷ lệ sai khác mong đợi} = t \times \frac{Cv}{\sqrt{r}}$$

Trong đó: t: giá trị trong phép thử Student; Cv: hệ số biến động; và r: số lần lặp lại. Giá trị Cv phù hợp có thể tìm qua các thí nghiệm tương tự được công bố trong các bài báo khoa học. Nó có thể dao động trong khoảng 5-25% đối với các thí nghiệm dạng này.

Quy mô thực tế của thí nghiệm thay đổi phụ thuộc cả số lần lặp lại và số nghiệm thức. Số lần lặp lại trong mỗi nghiệm thức có thể ít hơn đối với các thí nghiệm nhiều nhân tố (multi-factorial design) vì các thí nghiệm này có tổng số mẫu lớn hơn.

Kết luận được rút ra từ kết quả của một lần thí nghiệm nghiên cứu có thể không chính xác. Để có thể có kết luận đạt mức độ tin cậy cao hơn, nhiều khi thí nghiệm được làm lại trong điều kiện tương tự. Trường hợp đó được gọi là nhắc lại thí nghiệm. Nhắc lại thí nghiệm cũng là để đảm bảo rằng nếu nghiên cứu đó được thực hiện lại bởi người khác thì vẫn thu được kết quả tương tự.

e. Phân lô thí nghiệm

Khi đã chọn được các cá thể đại diện cho đối tượng nghiên cứu và đã quyết định được số nghiệm thức/lô thì việc tiếp theo là phân chia các cá thể động vật thí nghiệm vào các lô. Nguyên tắc phân lô là đảm bảo tính “đồng đều” giữa các lô, do đó mà các phương pháp ngẫu nhiên thường được áp dụng cho việc phân lô.

Tuy nhiên, trong một số trường hợp không thể chọn được toàn bộ các cá thể thí nghiệm “đồng đều” nhau thì các yếu tố khác nhau giữa các cá thể có thể làm “nhiều” kết quả thí nghiệm. Ví dụ, khi ta làm thí nghiệm theo dõi ảnh hưởng của một vài loại thức

ăn đến tốc độ tăng khối lượng của lợn con sau cai sữa; tuy nhiên, đàn lợn con thí nghiệm phải lấy từ các lợn mẹ khác nhau, ở các lứa đẻ, thời gian đẻ và khối lượng khác nhau... Tất cả những yếu tố “khác nhau” này đều có thể chi phối làm “nhiều” tốc độ tăng khối lượng của lợn khi thí nghiệm mà không chỉ có thức ăn thí nghiệm mới ảnh hưởng. Trong những trường hợp như thế, cần phải tính trước xem có thể dùng được các phương pháp thống kê phù hợp cho phép “loại trừ” được các yếu tố “nhiều” này hay không. Sau đây là hai phương pháp thường được áp dụng:

- Tạo khối (blocking)

Tạo khối là cách “cực chẳng đã” để xử lý các nguồn gây biến động không mong muốn nhưng không loại trừ được khi bố trí thí nghiệm. Tạo khối là phải tạo ra một biến độc lập thuộc dạng biến phân loại, đó có thể là các lứa đẻ khác nhau, các trang trại khác nhau, các vị trí thí nghiệm khác nhau ... Mỗi khối (block) phải có tất cả các lô/mức thí nghiệm (treatments). Đưa “khối” vào phân tích thống kê sẽ giúp loại trừ được ảnh hưởng của biến độc lập này ra khỏi sai số (error variation), dù nó làm mất đi một số độ tự do (= số khối – 1) cho sai số ngẫu nhiên. Ví dụ, để xử lý khối trong phân tích MINITAB khai lệnh như sau: *GLM output = block + treatment*. Không nên đưa tương tác giữa khối và lô trong mô hình phân tích thống kê vì điều này sẽ khó diễn giải kết quả, cho dù nó có ý nghĩa thống kê đi nữa.

- Sử dụng hiệp biến (covariate)

Sử dụng hiệp biến là một cách khác để loại trừ một nguồn biến động đã biết có ảnh hưởng (gây nhiễu) đến kết quả theo dõi các chỉ tiêu quan sát (biến phụ thuộc) của thí nghiệm. Hiệp biến là những biến liên tục như khối lượng ban đầu của gia súc thí nghiệm, năng suất sữa cá thể lúc bắt đầu thí nghiệm ... mà khi chọn mẫu thí nghiệm không thể chọn được “đồng đều” giữa các cá thể. Việc sử dụng các số liệu này làm hiệp biến là rất cần thiết trong các thí nghiệm chăn nuôi vì gia súc ở độ tuổi khác nhau hay ở giai đoạn tiết sữa khác nhau sẽ có phản ứng không giống nhau với yếu tố thí nghiệm.

Trong các thí nghiệm phân lô có hiệp biến, phải phân các cá thể vào các lô sao cho giá trị trung bình và độ biến động giá trị của hiệp biến của các cá thể là tương đương nhau giữa các lô. Ví dụ, nếu ta phải dùng khối lượng ban đầu như một hiệp biến (vì không đồng đều giữa các cá thể), mặc dù ta không phân lô hay tạo khối theo khối lượng, nhưng phải phân bổ các cá thể có khối lượng khác nhau vào các lô sao cho khối lượng bình quân và độ lệch chuẩn về khối lượng ban đầu phải tương đương nhau giữa các lô. Chú ý: Khi đưa một yếu tố nào đó làm hiệp biến trong mô hình phân tích thống kê thì sẽ mất đi một độ tự do của sai số ngẫu nhiên.

1.2.5. Phân tích số liệu thí nghiệm

a. Kiểm tra và xử lý số liệu

Thu thập số liệu thí nghiệm hàng ngày, hàng tuần hay hàng tháng sẽ có được kết quả như lượng thu nhận thức ăn, khối lượng cơ thể qua các thời gian thí nghiệm,... Tuy nhiên, từ các chỉ tiêu này có thể cần phải tính toán thêm một số chỉ tiêu quan tâm khác

như tăng khối lượng bình quân hàng ngày, hệ số chuyển hóa thức ăn,... cho từng cá thể/đơn vị thí nghiệm. Hơn nữa, các số liệu phải được sắp xếp lại dưới dạng các tệp (file) dữ liệu thống kê, trong đó phải thể hiện được các biến trong mô hình phân tích (xem mục 1.2.5b). Đối với những dữ liệu ban đầu và việc tính toán sơ bộ như vậy thì nên lưu giữ và thực hiện trên các phần mềm quản lý dữ liệu đơn giản như Excel, mặc dù có thể quản lý trực tiếp trên các phần mềm thống kê. Những bộ số liệu đó có thể dễ dàng chuyển (import) vào phần mềm thống kê khi cần phân tích thống kê.

Khi đã có bộ số liệu cuối cùng (số liệu gốc) thì vấn đề quan trọng tiếp theo là rà soát kiểm tra tính chính xác của số liệu, phát hiện những số liệu trống (missing data), số liệu không bình thường. Các số liệu xuất hiện không bình thường, trái với dự kiến ban đầu hay quy luật chung có thể là do sai sót trong quá trình đo lường hay ghi chép, nhưng cũng có thể vì những lý do đặc biệt khác mà người nghiên cứu không nên bỏ qua. Đôi lúc đó lại là đầu mối cho những phát hiện tiềm tàng hay gợi ý cho những nghiên cứu tiếp theo. Mặt khác, cần phải kiểm tra xem các số liệu thu thập được có thỏa mãn các điều kiện để sử dụng mô hình phân tích thống kê dự kiến hay không. Ví dụ, nếu muốn phân tích phương sai thì trước hết số liệu phải tuân theo phân phối chuẩn.

b. Mô hình phân tích thống kê

Phân tích thống kê số liệu có vai trò rất quan trọng, cho phép rút ra được những kết luận có ý nghĩa từ thí nghiệm. Do đó chọn *mô hình phân tích thống kê* phải phù hợp với số liệu thu thập và phân tích được thông qua một phần mềm máy tính. Trên thực tế thì mô hình phân tích thống kê thường đã được dự kiến khi thiết kế (trước khi tiến hành) thí nghiệm. Tuy nhiên, sau khi đã có số liệu của thí nghiệm thì cần kiểm tra lại tính phù hợp của mô hình phân tích với số liệu thu được.

Ví dụ: Một mô hình thống kê có thể mô tả như sau:

$$Y_j = \mu + \sum_{i=1}^n \alpha_i X_i + \varepsilon_j$$

Trong đó: Y_j : giá trị của biến phụ thuộc ở quan sát j

μ : giá trị trung bình chung

α_i : ảnh hưởng của biến độc lập thứ i

X_i : bộ các biến độc lập ($i = 1, \dots, n$)

ε_j : sai số ngẫu nhiên; các ε_j độc lập, theo phân phối chuẩn $N(0, \sigma^2)$.

Tương tự như trong giả thuyết khoa học, mỗi mô hình phân tích thống kê phải có hai loại biến số: *biến độc lập* và *biến phụ thuộc*. Ví dụ trong phương trình hồi quy tuyến tính bậc nhất $Y = aX + b$ có X là biến độc lập và Y là biến phụ thuộc (Y phụ thuộc vào X). Biến phụ thuộc thường là các chỉ tiêu theo dõi thí nghiệm (như khối lượng cơ thể, tăng khối lượng, hệ số chuyển hóa thức ăn...) và còn được gọi là biến đáp ứng (responsive variable). Biến động giá trị của các chỉ tiêu (biến) này thường “phụ thuộc” vào các yếu tố thí nghiệm (như phẩm giống, loại thức ăn ...) hay một số yếu tố khác mà

không loại trừ được khi bố trí thí nghiệm (như khối lượng ban đầu, điều kiện thời tiết...). Chính vì thế, những yếu tố chi phối (biến độc lập) này còn được gọi là biến giải thích (explanatory variables). Một thí nghiệm có thể có một hoặc nhiều yếu tố thí nghiệm (biến độc lập) và các yếu tố thí nghiệm này có thể là yếu tố cố định hoặc yếu tố ngẫu nhiên. Tuy nhiên, trong mỗi mô hình thống kê chỉ có một biến phụ thuộc (tương ứng với 1 chỉ tiêu theo dõi thí nghiệm) và 1 hay nhiều biến độc lập/giải thích.

Biến độc lập là các yếu tố, điều kiện mà khi thay đổi trên đối tượng nghiên cứu sẽ ảnh hưởng đến kết quả của các chỉ tiêu theo dõi thí nghiệm (tức là biến phụ thuộc). Trong mỗi biến độc lập, thường có một mức đối chứng (chứa các yếu tố, điều kiện ở mức độ thông thường hoặc đã được xác định mà người nghiên cứu không cần tiên đoán ảnh hưởng của chúng). Các mức còn lại sẽ được so sánh với nghiệm thức đối chứng hoặc so sánh giữa các cặp với nhau. Ví dụ, trong nghiên cứu “*Ảnh hưởng hàm lượng protein trong khẩu phần đến tốc độ tăng khối lượng của thỏ New Zealand sinh trưởng*” có các biến như sau:

+ *Biến độc lập*: hàm lượng protein thô trong khẩu phần với các nghiệm thức trong thí nghiệm có thể chứa 12, 14, 16, 18 và 20% protein trong khẩu phần trong vật chất khô, trong đó nghiệm thức “đối chứng” chứa 16% là mức protein thô được khuyến cáo sử dụng trong sản xuất.

+ *Biến phụ thuộc*: tốc độ tăng khối lượng của thỏ (g/con/ngày).

Như vậy, về mặt thống kê mô hình phân tích là mô hình “nhân - quả” (cause and effect) với biến độc lập là “nhân” còn biến phụ thuộc là “quả”. Tuy nhiên trong thực tế nhiều lúc khó xác định biến nào phụ thuộc biến nào. Do đó, người ta thường xác định các biến trong thí nghiệm dựa trên mối quan hệ “nhân - quả” của giả thuyết. Kết quả quan sát lệ thuộc vào nguyên nhân gây ảnh hưởng. Dựa vào mối quan hệ trong giả thuyết đặt ra, người nghiên cứu dễ dàng xác định được yếu tố (factor) nào ảnh hưởng đến chỉ tiêu quan sát (observed parameter). Ví dụ, “Ảnh hưởng của khoảng cách lứa cắt đến tốc độ tái sinh và năng suất của cỏ voi”. Ở đây, tốc độ tái sinh và năng suất (kết quả) khác nhau là do ảnh hưởng của khoảng cách lứa cắt (nguyên nhân) khác nhau. Nói một cách khác, biến độc lập là yếu tố mà người nghiên cứu có ý định làm thay đổi (khoảng cách lứa cắt) và biến phụ thuộc là biến chịu ảnh hưởng của các yếu tố thay đổi đó (ở đây là tốc độ tái sinh và năng suất của cỏ voi).

Trong nghiên cứu chăn nuôi cũng như sinh học nói chung, quan hệ “nhân - quả” trong mô hình thống kê như trên nhiều khi chỉ mang tính ước lệ. Trong nhiều trường hợp biến phụ thuộc thực ra không phụ thuộc vào biến độc lập về mặt bản chất, mặc dù khi phân tích thì sự “phụ thuộc” đó lại có thể “có ý nghĩa thống kê”. Do vậy, khi giải nghĩa mối quan hệ giữa các biến cần hết sức thận trọng, không phải lúc nào biến độc lập cũng là “nguyên nhân” gây ra sự biến đổi của biến “phụ thuộc”, mà có thể cả hai đều cùng là “quả” của một “nhân” khác nên chúng có “quan hệ” với nhau. Ví dụ, nếu chúng ta phân tích hồi quy giữa hàm lượng mỡ và hàm lượng protein trong sữa bò thì sẽ thấy chúng có tương quan chặt với nhau, nhưng đó không phải là quan hệ “nhân - quả” và có thể dùng chỉ tiêu nào làm biến độc lập hay biến phụ thuộc cũng được. Trong những mối quan hệ

kiểu như vậy chúng ta không nên vội rút ra kết luận “nhân - quả” mỗi khi phân tích thống kê cho thấy “có ý nghĩa” mà phải tìm hiểu cơ chế và bản chất thực sự của mối quan hệ.

Về mặt giá trị, nhiều số liệu thí nghiệm chăn nuôi thuộc dạng *biến liên tục* (continuous variable) như khối lượng cơ thể, lượng thức ăn thu nhận,... Kiểu biến động số liệu của những biến này tuân theo “phân phối chuẩn”. Một số chỉ tiêu nhất định lại có số liệu thuộc dạng “có/không” như “thụ thai/không thụ thai” và thuộc loại *biến rời rạc* có quy luật biến động theo phân phối nhị phân. Biến rời rạc (discrete variables) còn có thể là những biến có số liệu thuộc dạng “loại 1, loại 2, loại 3...”, nên được gọi là *biến phân loại* (categorical variables), chúng tuân theo phân phối Poisson. Tuy nhiên, trong nhiều trường hợp những số liệu dạng như vậy có thể được “chuyển dạng” hay “đổi biến” bằng các công cụ thống kê (như logarit hoá, khai căn bậc hai,...) để làm cho chúng tuân theo quy luật phân phối chuẩn. Số liệu theo kiểu tỷ lệ phần trăm cũng có thể chuyển dạng. Tuy nhiên cũng có trường hợp chúng ta không thể hoán chuyển số liệu bằng các hàm thông dụng như logarit hay hàm mũ.

Tùy theo dạng biến phụ thuộc mà có thể áp dụng các phương pháp phân tích thống kê khác nhau. Biến phụ thuộc có số liệu liên tục có thể phân tích thống kê bằng một loạt công cụ được gọi là *thống kê tham số* (parametric statistics), gồm phân tích phương sai (Analysis of Variance - ANOVA), hiệp phương sai (Analysis of Covariance - ANCOVA) và phân tích hồi quy (Regression Analysis - REGRESS), các phân tích thống kê nhiều chiều như xếp loại (Classification), phân biệt (Discriminant), chia chùm (Cluster analysis), thành phần chính (Principal component) và cũng có thể áp dụng thống kê phi tham số. Phân tích phương sai áp dụng cho các mô hình thống kê có biến phụ thuộc là biến liên tục còn biến độc lập là các biến phân loại (categorical variables) như phẩm giống vật nuôi, loại khẩu phần ăn... Phân tích hồi quy áp dụng cho mô hình thống kê mà cả biến độc lập và biến phụ thuộc đều là biến liên tục. Hiện nay, các phần mềm thống kê mạnh có thể cho phép phân tích các mô hình thống kê mà trong đó các biến độc lập có thể bao gồm cả biến phân loại và biến liên tục thông qua mô hình tuyến tính tổng quát (Generalized Linear Model - GLM). Biến phụ thuộc loại rời rạc có thể phân tích phương pháp hồi quy logistic, phép thử χ^2 hoặc phép thử chính xác của Fisher. Thống kê phi tham số (non-parametric statistics) được áp dụng khi cách tiếp cận thống kê tham số không phù hợp.

Chương 2 sẽ hướng dẫn cách sử dụng các mô hình thống kê tham số và phi tham số để phân tích số liệu thí nghiệm thông qua các ví dụ cụ thể trên phần mềm SAS (Statistical Analysis System).

1.3. ĐỀ CƯƠNG NGHIÊN CỨU KHOA HỌC

1.3.1. Đề cương nghiên cứu là gì?

Từ ý tưởng nghiên cứu cho một đề tài khoa học (vấn đề/câu hỏi nghiên cứu, giả thuyết khoa học), một kế hoạch mô tả chi tiết công trình dự định nghiên cứu phải được xây dựng và thể hiện một cách có hệ thống trong một đề cương nghiên cứu (Research

Proposal) hay Thuyết minh đề tài. Đây là một tài liệu quan trọng về một đề tài/công trình nghiên cứu, có thể ví như bản vẽ thiết kế của một công trình xây dựng. Nó phải được viết sao cho có thể thuyết phục người khác rằng đề tài nghiên cứu đó là quan trọng và đáng thực hiện, đồng thời tác giả có khả năng thiết kế và thực hiện đề tài nghiên cứu để đạt được mục tiêu đề ra. Trên cơ sở đó hội đồng khoa học/cơ quan tài trợ có thể xem xét và duyệt/cấp kinh phí cho nghiên cứu.

Viết đề cương nghiên cứu vì thế vừa mang tính tiếp thị, vừa là khoa học vừa là nghệ thuật. Cách viết đề cương nghiên cứu khác với viết bài báo khoa học (xem Chương 3) về mục tiêu và thời gian tính. Mục tiêu của viết Đề cương nghiên cứu là thuyết phục nhà tài trợ/Hội đồng xét duyệt đề tài rằng ý tưởng nghiên cứu là tốt, có cách tiếp cận, cách giải quyết vấn đề hợp lý, điều kiện nghiên cứu là khả thi. Mục tiêu của viết bài báo khoa học là để công bố những phát hiện cụ thể từ một công trình nghiên cứu. Về thời gian tính, viết đề cương nghiên cứu là viết về những việc sẽ làm trong tương lai, còn viết bài báo khoa học là viết về những gì đã làm trong quá khứ.

1.3.2. Nội dung của đề cương nghiên cứu

Nội dung của một đề cương nghiên cứu tùy thuộc vào yêu cầu của cơ quan tài trợ/quản lý khoa học. Tuy nhiên, yêu cầu tối thiểu trong các đề cương nghiên cứu phải có các nội dung sau:

- Tên đề tài
- Tổng quan và Tính cấp thiết
- Giả thuyết khoa học
- Mục tiêu của đề tài
- Nội dung nghiên cứu
- Phương pháp nghiên cứu
- Các nguồn lực dự kiến
- Kết quả dự kiến
- Ý nghĩa và tác động
- Kế hoạch triển khai
- Kinh phí
- Tóm tắt

Mỗi cơ quan tài trợ/quản lý nghiên cứu đều có biểu mẫu riêng cho đề cương nghiên cứu/thuyết minh đề tài. Các nhà khoa học phải tuân thủ theo đó khi viết để đề nghị xét duyệt và cấp kinh phí nghiên cứu. Không đáp ứng được thể thức theo quy định của nhà tài trợ là một trong những lý do phổ biến nhất đề cương nghiên cứu không được chấp nhận.

1.3.3. Chi tiết các phần của đề cương nghiên cứu

a. Tên đề tài

Tên đề tài là ấn tượng đầu tiên của người đọc, cần viết ngắn gọn, súc tích, rõ ràng để cho người đọc hiểu ngay được các ý đồ của công trình nghiên cứu.

Một tên đề tài tốt thường nêu lên được một câu hỏi hay chứa đựng một tương phản rõ ràng có thể làm cho người đọc dễ nhớ hơn và tạo cho người đọc một cảm giác chung về điều gì tác giả muốn nghiên cứu.

Một tên đề tài quá dài hay quá sâu về kỹ thuật sẽ không thu hút được sự chú ý và quan tâm của người đọc (người duyệt Đề cương nghiên cứu). Ngược lại tên đề tài quá ngắn và chung chung làm cho người đọc cân nhắc khi xem xét cấp kinh phí vì không hiểu rõ ý đồ nghiên cứu.

b. Tổng quan và tầm quan trọng của đề tài

Phần này dùng để đặt công trình nghiên cứu vào một bối cảnh: những gì đã được làm trước đây? Bằng cách nào công trình nghiên cứu đề xuất bổ sung thêm về lý luận và thực tiễn? Đây là yếu tố mới trong công trình nghiên cứu này? Đồng thời phần này cũng nhằm chứng minh sự hiểu biết và năng lực của người đề xuất trong lĩnh vực nghiên cứu.

Trong phần *Tổng quan tài liệu/Bối cảnh* và *Tính cấp thiết/Tầm quan trọng của đề tài* tác giả cần phác họa được một bức tranh tổng quan về quá khứ và hiện tại của chủ đề nghiên cứu (đã biết), tìm cho ra được khoảng trống tri thức (chưa biết/chưa rõ) để đặt ra câu hỏi nghiên cứu làm cơ sở cho từng mục tiêu cụ thể của đề tài. Như vậy, mỗi mục tiêu cụ thể đặt ra cho đề tài đều phải được lý giải trong phần *Bối cảnh*.

Phần này phải được dẫn dắt một cách khoa học, nhưng không quá phức tạp sao cho những người không cùng chuyên ngành sâu và đặc biệt là những người xét duyệt đề tài có thể hiểu được vấn đề và bị thuyết phục. Trước hết tác giả phải điễm qua được những bài báo quan trọng trước đây và những công trình nghiên cứu mới nhất có liên quan đến từng câu hỏi nghiên cứu. Khi điễm qua như vậy cần chỉ ra được những thông tin cụ thể, làm nổi bật được những khía cạnh tích cực cũng như những hạn chế của các công trình nghiên cứu trước. Nêu rõ những quan điễm khác nhau và những lĩnh vực còn tranh luận. Tránh dùng những kiến thức phổ biến từ giáo trình hay liệt kê tràn lan các trích dẫn tài liệu tham khảo mà không sắp xếp theo một ý đồ nào cả. Cấu trúc Tổng quan tài liệu nên đi từ các thông tin rộng và cụ thể hóa dần, hướng Tổng quan tới một hay một vài câu hỏi cụ thể chưa có câu trả lời.

Trong *Thuyết minh đề tài* khoa học và công nghệ xin tài trợ trong nước, thì ngoài việc phân tích, đánh giá được những công trình nghiên cứu trên thế giới có liên quan và những kết quả nghiên cứu mới nhất, những bước tiến về trình độ khoa học và công nghệ của những kết quả nghiên cứu đó thì tác giả còn phải phân tích, đánh giá tình hình nghiên cứu trong nước thuộc lĩnh vực nghiên cứu của đề tài, đặc biệt phải nêu cụ thể được những kết quả khoa học và công nghệ liên quan đến đề tài mà các cán bộ tham gia đề tài đã thực hiện. Nếu có các đề tài cùng bản chất đã và đang được thực hiện ở cấp khác, nơi khác thì phải giải trình rõ các nội dung kỹ thuật liên quan đến đề tài đang đề

ngợi. Trong phần này tác giả cũng có thể tóm tắt những kết quả nghiên cứu ban đầu của mình hay của nhóm nghiên cứu liên quan đến vấn đề nghiên cứu để thuyết phục người đọc rằng tác giả đã có kiến thức, kinh nghiệm, đã nắm được phương pháp để thực hiện tốt công trình nghiên cứu đang đề nghị.

Trên cơ sở đánh giá tình hình nghiên cứu trong và ngoài nước, phân tích những công trình nghiên cứu có liên quan, những kết quả mới nhất trong lĩnh vực nghiên cứu đề tài, đánh giá những khác biệt về trình độ khoa học và công nghệ trong nước và thế giới, những vấn đề đã được giải quyết, cần nêu rõ những vấn đề còn tồn tại, chỉ ra những hạn chế cụ thể, tìm ra khoảng trống trong tri thức hiện tại về chủ đề nghiên cứu, từ đó nêu được hướng giải quyết vấn đề, luận giải cho mục tiêu đặt ra của đề tài cũng như những nội dung cần thực hiện để đạt được mục tiêu.

Tóm lại, tác giả phải chứng minh được nghiên cứu mình đề xuất là dựa trên và là sự tiếp nối những công trình trước đây trong một lĩnh vực cụ thể, nhưng sẽ là một sự phát triển quan trọng trong lĩnh vực đó. Đồng thời cần phác thảo được tương lai của chủ đề nghiên cứu, giải thích tại sao chủ đề nghiên cứu là quan trọng để thuyết phục người đọc rằng công trình nghiên cứu là cần thiết, sẽ có những đóng góp mới về lý luận và thực tiễn, chính vì thế mà công trình nghiên cứu đáng để đầu tư. Phần này cũng nhằm cung cấp cho người đọc những thông tin cần thiết để hiểu được mục tiêu và cách tiếp cận của công trình nghiên cứu.

c. Giả thuyết khoa học

Trong nhiều mẫu *Đề cương nghiên cứu* thường có phần/mục dành để tác giả nêu giả thuyết (một hay nhiều) cho công trình nghiên cứu đề xuất. Ngoài việc nêu giả thuyết như đã trình bày trong mục 1.2.3 tác giả cần nêu được lý do hay tính cấp thiết tại sao (những) giả thuyết đó lại quan trọng và cần tìm hiểu.

d. Mục tiêu nghiên cứu

Điều quan trọng nhất để có được một nghiên cứu tốt là việc xác định mục tiêu nghiên cứu. Cho dù thiết kế thí nghiệm có tốt đến đâu, phương pháp nghiên cứu có tinh vi đến đâu, cách xử lý thống kê kết quả thí nghiệm có chuẩn xác đến đâu đi nữa thì công trình nghiên cứu cũng ít có giá trị nếu như không trả lời được một câu hỏi có tầm quan trọng về mặt khoa học và thực tiễn hay cả hai. Mục tiêu của một nghiên cứu là để để giải quyết một vấn đề nghiên cứu hay trả lời cho một hay một vài câu hỏi nghiên cứu. Do vậy mục tiêu phải phù hợp với tổng quan, tính cấp thiết, câu hỏi nghiên cứu và giả thuyết, đồng thời nó phải thể hiện được cái mới của đề xuất nghiên cứu.

Khi viết phần *Mục tiêu nghiên cứu* nên chia thành 2 phần:

- *Mục tiêu chung* hay *Mục tiêu tổng quát* là mục tiêu lâu dài mà nhà nghiên cứu muốn đạt được nhằm giải quyết một vấn đề nghiên cứu. Có thể đề tài nghiên cứu hiện tại chưa giải quyết được hết, nhưng nó sẽ đóng góp một phần trong định hướng giải quyết vấn đề đó về lâu về dài.

Đối với ngành chăn nuôi nước ta mục tiêu chung của các nghiên cứu khoa học chăn nuôi phải nhằm hướng tới tầm nhìn xây dựng một ngành “*Chăn nuôi văn minh/thông*

minh” bền vững phù hợp với xu hướng phát triển chăn nuôi của thế giới. “*Chăn nuôi văn minh*” là một hoạt động kinh tế thông qua sản xuất các sản phẩm có nguồn gốc động vật nhưng phải tôn trọng lợi ích của các đối tượng liên quan là: (1) Người chăn nuôi (an toàn và hiệu quả kinh tế), (2) Người tiêu dùng (chất lượng và vệ sinh an toàn thực phẩm), (3) Môi trường sinh thái (giảm thiểu ô nhiễm và bảo tồn đa dạng sinh học), và (4) Vật nuôi (đảm bảo phúc lợi động vật/animal welfare). Mục tiêu chung của một nghiên cứu chăn nuôi cần hướng tới ít nhất là một trong những mục tiêu tổng quát này thông qua việc giải quyết các vấn đề của thực tiễn chăn nuôi đặt ra. Đó có thể là những vấn đề thuộc về lĩnh vực giống vật nuôi, thức ăn chăn nuôi, tổ chức sản xuất và ngành hàng, quản lý sức khỏe và sinh sản, xử lý chất thải và giảm thiểu ô nhiễm môi trường, ...

- *Mục tiêu cụ thể* (specific objectives) là những điểm cần đạt được thông qua thực hiện đề tài nghiên cứu nhằm hướng tới *Mục tiêu chung*. Thông thường mỗi công trình nghiên cứu nên có 2-3 mục tiêu cụ thể không phụ thuộc lẫn nhau. Mỗi mục tiêu cụ thể thường đi kèm với một giả thuyết khoa học để trả lời cho một câu hỏi nghiên cứu. Mục tiêu cụ thể là nền tảng cho các nội dung hoạt động của đề tài, làm cơ sở cho việc đánh giá kế hoạch nghiên cứu đã đưa ra và là điều mà kết quả nghiên cứu phải đạt được.

Mục tiêu cụ thể cho một nghiên cứu chăn nuôi cần phải thể hiện được các yêu cầu SMART sau đây:

- + Cụ thể cho từng đối tượng vật nuôi (**Specific**);
- + Có khả năng đo lường, kiểm chứng được (**Measurable**);
- + Trên cơ sở các nguồn lực sẵn có (**Available resources**);
- + Phù hợp với thực tiễn chăn nuôi (**Realistic**);
- + Thực hiện được trong một thời gian nhất định (**Time bound**).

Do vậy, việc nghiên cứu tài liệu, suy nghĩ về các câu hỏi đặt ra và thảo luận với đồng nghiệp, đặc biệt là với người chăn nuôi, là đối tượng sẽ ứng dụng kết quả nghiên cứu, là phần quan trọng nhất trong việc xác định các mục tiêu cụ thể cho một nghiên cứu chăn nuôi. Các mục tiêu cụ thể nên tương đối độc lập với nhau nhưng đều góp phần để đạt được mục tiêu chung.

Một thủ thuật có thể sử dụng để xây dựng mục tiêu nghiên cứu là đi từ “cây vấn đề”. Một vấn đề cốt lõi cần quan tâm nghiên cứu có các nguyên nhân hay các yếu tố cấu thành nên nó theo giả thuyết của người nghiên cứu (xây dựng cây vấn đề). Từ cây vấn đề chuyển thành “cây mục tiêu”: chuyển vấn đề cốt lõi thành mục tiêu chung; chuyển các nguyên nhân hay các yếu tố cấu thành của vấn đề thành các mục tiêu cụ thể. Ví dụ, một vấn đề nổi cộm trong ngành chăn nuôi bò sữa của nước ta hiện nay là năng suất sinh sản của bò cái Holstein Friesian thuần rất thấp so với các loại bò lai nước hay so với chính bò Holstein Friesian nuôi ở các nước ôn đới. Câu hỏi đặt ra là đâu là nguyên nhân của hiện tượng này. Có thể giả thuyết rằng nguyên nhân làm cho năng suất sinh của bò cái Holstein Friesian thuần nuôi ở nước ta thấp là do: (1) stress nhiệt, (2) chất lượng thức ăn thấp và/hay khẩu phần mất cân đối dinh dưỡng (có thể có nhiều nguyên nhân khác). Điều này đòi hỏi phải làm thí nghiệm mới trả lời chính xác được. Do đó, mục tiêu chung của nghiên cứu là nâng

cao năng suất sinh sản của bò cái Holstein Friesian thuần (chuyển ngược từ năng suất sinh sản thấp) và các mục tiêu cụ thể (khắc phục các nguyên nhân theo giả thuyết) của nghiên cứu là đánh giá ảnh hưởng của (1) việc làm mát (giảm stress nhiệt) và (2) cân bằng dinh dưỡng khẩu phần đến năng suất sinh sản của bò cái Holstein Friesian thuần nuôi ở nước ta.

e. Phương pháp nghiên cứu

Đây thường là phần được viết dài nhất trong *Đề cương nghiên cứu* để trả lời càng rõ ràng, càng thuyết phục càng tốt câu hỏi là nhà nghiên cứu sẽ làm gì và làm như thế nào để đạt được mục tiêu của đề tài. Trong phần này, trước hết nhà nghiên cứu phải đề ra được các nội dung nghiên cứu, tức là cần làm gì để thực hiện được từng mục tiêu cụ thể của đề tài trên cơ sở các giả thuyết khoa học đã được đặt ra cho nghiên cứu. Có thể phải thực hiện nhiều nội dung nghiên cứu mới đạt được một mục tiêu. Trong *Đề cương nghiên cứu* nhiều khi *Nội dung nghiên cứu* được viết trong một mục riêng đi trước mục *Phương pháp nghiên cứu* cụ thể, nhất là khi một phương pháp nghiên cứu cụ thể được áp dụng cho nhiều nội dung khác nhau.

Một số mẫu *Đề cương nghiên cứu* yêu cầu đưa thêm mục *Phương pháp tiếp cận* (*Approach*) để trả lời câu hỏi bằng cách nào để đạt được mục tiêu. Thực ra, phương pháp tiếp cận cũng thuộc về phương pháp, nhưng đó là tuyên bố định hướng về cách thức nghiên cứu để làm khung giới hạn cho việc lựa chọn các phương pháp nghiên cứu cụ thể. Để dễ hiểu, giả sử mục tiêu của chúng ta là đi từ Hà Nội đến Thành phố Hồ Chí Minh; tùy theo hoàn cảnh, chúng ta có thể chọn đi bằng cách nào trong một số cách đi (cách tiếp cận) khác nhau: đường hàng không, đường sắt, đường bộ hay đường thủy; và khi đã chọn một “đường” để đi thì các việc tiếp theo (như mua vé, lịch trình, phương tiện ra ga/bến...) hoàn toàn phụ thuộc vào phương thức/cách đi đã chọn đó.

Phương pháp nghiên cứu phải được thiết kế sao cho công trình nghiên cứu có thể cung cấp được kết quả nhằm trả lời được từng câu hỏi nghiên cứu đặt ra, đồng thời phải đảm bảo tính khả thi trong khuôn khổ các nguồn lực cho phép (gia súc, chuồng trại, trang bị phòng thí nghiệm, ...). Nói một cách khác, các phương pháp cụ thể phải phù hợp với từng nội dung nghiên cứu cụ thể nhằm đánh giá được ảnh hưởng của một hay nhiều yếu tố lên các chỉ tiêu cần quan sát. Hơn nữa, khi mô tả từng kỹ thuật sử dụng trong nghiên cứu cụ thể, trước hết cần nêu được mục đích, sau đó mô tả mô hình, công cụ, vật liệu sử dụng và quy trình thực hiện. Việc mô tả phải chi tiết để người khác có thể thực hiện được. Tác giả cũng cần giải thích tại sao lại chọn kỹ thuật, công cụ hay mô hình đó. Nếu tác giả có những thay đổi so với các tác giả trước thì cần nêu rõ đâu là những điểm tương đồng và đâu là những cải biên.

Thông thường phần này có những tiêu mục như:

- *Thiết kế thí nghiệm*: Mô tả thiết kế thí nghiệm hay phương pháp thu thập số liệu nghiên cứu.

- *Phương pháp chọn mẫu*: Đối tượng nghiên cứu (tiêu chuẩn mẫu, đối tượng loại trừ), phương pháp chọn mẫu (chi tiết từng phương pháp/tên phương pháp tham chiếu, những cải tiến về phương pháp), số lượng mẫu (lý do xác định cỡ mẫu, khả năng đảm bảo độ tin cậy của kết quả, có nhóm đối chứng hay đối sánh hay không?)

- *Địa điểm nghiên cứu*: Nơi tiến hành nghiên cứu, có sự ủng hộ của chính quyền địa phương không?

- *Thủ tục/Quy trình tiến hành thí nghiệm*: Chi tiết từng bước tiến hành, cách chọn mẫu và thu thập số liệu, mô tả chi tiết yếu tố thí nghiệm, công cụ đo lường được sử dụng (có cả độ chính xác và copy cả công cụ đó đưa vào trong bản đề xuất), khung thời gian tiến hành nghiên cứu.

- *Phân tích thống kê*: Mô hình phân tích thống kê nào sẽ được dùng để phân tích số liệu? Các mô hình phân tích thống kê phải phù hợp với thiết kế thí nghiệm.

Tóm lại, ứng với mỗi mục tiêu cụ thể của đề tài tác giả cần luận cứ rõ cách tiếp cận và thiết kế nội dung nghiên cứu để đạt được. Đối với mỗi nội dung nghiên cứu, cần có phương pháp nghiên cứu và kỹ thuật cụ thể sẽ sử dụng. Để tăng tính thuyết phục tác giả nên chứng minh tính khả thi của các phương pháp áp dụng và so sánh chúng với các phương pháp giải quyết tương tự khác, đồng thời phân tích để làm rõ được tính mới, tính độc đáo, tính sáng tạo của đề tài đang đề xuất. Việc mô tả phương pháp nghiên cứu phải chi tiết, chính xác và rõ ràng để người khác cũng có thể hiểu và thực hiện được việc nghiên cứu dựa trên mô tả đó.

g. Các nguồn lực dự kiến

Tác giả cần tóm tắt không gian, trang thiết bị, nhân lực và các nguồn lực cần thiết khác cần để thực hiện nghiên cứu. Đặc biệt về nhân lực, phải xác định trình độ cần thiết cho công trình nghiên cứu và sự phù hợp với công việc được bố trí cho mỗi người. Cần nêu rõ những người tham gia chính và những người khác sẽ có đóng góp quan trọng cho công trình nghiên cứu. Tóm tắt lý lịch các nhân sự chính: quá trình đào tạo, kinh nghiệm công tác, các công bố khoa học, các công trình nghiên cứu gần đây. Ngoài ra, nên bổ sung thư ủng hộ đối với việc chia sẻ các nguồn lực cần cho công trình nghiên cứu và giải trình khả năng huy động các nguồn ngoại lực cho nghiên cứu.

h. Kết quả dự kiến

Mô tả những đóng góp về kiến thức và tầm quan trọng cho nghiên cứu trong tương lai, những lợi ích đối với các đối tượng hưởng lợi. Kết quả dự kiến phải phù hợp với mục tiêu và các nội dung nghiên cứu. Đồng thời cần nêu rõ cách thức phổ biến kết quả nghiên cứu (công bố bài báo, hội nghị/hội thảo, thương mại hóa, website...).

i. Ý nghĩa và tác động

Mục này là cơ hội cuối cùng để tác giả gây ấn tượng cho người đọc về tầm quan trọng của của công trình đang đề xuất và cung cấp ý tưởng cho định hướng tương lai của nghiên cứu. Mục này phải lý giải được tại sao nghiên cứu này có tính sáng tạo và kết quả nghiên cứu sẽ có đóng góp gì về mặt khoa học và thực tiễn.

k. Kế hoạch triển khai

Nêu rõ tổng thời gian cần cho công trình nghiên cứu và các mốc thời gian quan trọng cho các hoạt động chính. Ví dụ:

Tháng 1-3: Chuẩn bị

Tháng 4-10: Thu thập số liệu

Tháng 11-12: Phân tích số liệu

l. Kinh phí

Kinh phí ước tính phải phù hợp với khối lượng công việc và tính phức tạp của các hoạt động. Các cơ quan tài trợ có chính sách khác nhau khi tài trợ cho các chi phí trực tiếp (như nhân công và thiết bị) và các chi phí gián tiếp (như phí quản lý cơ quan). Cơ quan tài trợ có thể yêu cầu chứng minh trước kinh phí đối ứng của cơ quan chủ trì và tham gia nghiên cứu.

m. Tóm tắt

Tóm tắt đề tài/dự án phải chứa các thông tin đã được nêu trong các mục ở trên một cách ngắn gọn, rõ ràng, súc tích và logic với nhau. Cần có tóm tắt tổng quan dẫn đến khoảng trống trong tri thức hiện tại để làm cơ sở cho câu hỏi và giả thuyết nghiên cứu. Mục tiêu, nội dung, phương pháp nghiên cứu và kết quả dự kiến phải được trình bày ăn khớp với nhau trong một chỉnh thể thống nhất. Ngoài ra, cũng cần nêu rõ tác động, ý nghĩa và tính mới của đề xuất nghiên cứu. Nói cách khác, mục này cần trả lời một cách ngắn gọn và rõ ràng nhất nhưng đủ để người đọc biết được tác giả dự định làm gì (what?), tại sao (why?), để đạt được cái gì (what for?), bằng cách nào (how?), ở đâu (where?), lúc nào (when?), ai làm (who?) và tốn kém bao nhiêu (how much/many?).

CÂU HỎI ÔN TẬP

1. Tri thức khoa học là gì?
2. Mục tiêu và bản chất của nghiên cứu khoa học là gì?
3. Thế nào là phương pháp nghiên cứu khoa học?
4. Quy trình nghiên cứu chăn nuôi gồm những bước nào? Mối liên hệ giữa các bước?
5. Vấn đề và câu hỏi nghiên cứu là gì?
6. Ý tưởng nghiên cứu trong chăn nuôi có thể có từ đâu?
7. Giả thuyết khoa học là gì? Phương pháp nêu giả thuyết khoa học?
8. Mục đích và phương pháp của việc tìm hiểu cơ sở lý luận và thực tiễn cho đề tài nghiên cứu khoa học?
9. Những yếu tố cơ bản trong thí nghiệm chăn nuôi là gì? Nêu cụ thể từng yếu tố?
10. Thế nào là một mô hình phân tích thống kê? Các thành phần của nó?
11. Nêu tổng thể các phương pháp phân tích thống kê số liệu thí nghiệm chăn nuôi?
12. Đề cương nghiên cứu là gì?
13. Nội dung chính của Đề cương nghiên cứu bao gồm những phần nào?
14. Nêu cách viết từng phần của Đề cương nghiên cứu ?

Chương 2

PHÂN TÍCH SỐ LIỆU THÍ NGHIỆM BẰNG PHẦN MỀM MÁY TÍNH

Có nhiều phần mềm máy tính khác nhau có thể dùng để phân tích thống kê số liệu thí nghiệm chăn nuôi nhưng phổ biến nhất hiện nay là SAS (Statistical Analysis System). Chương này không nhắc lại nguyên lý thống kê của các mô hình thiết kế thí nghiệm mà chỉ hướng dẫn ứng dụng các mô hình thống kê phổ biến để phân tích số liệu của các thí nghiệm chăn nuôi thông qua phần mềm SAS để minh họa. Trong mỗi ví dụ sẽ có dữ liệu nhập, mô hình phân tích, các câu lệnh, kết quả phân tích thống kê của phần mềm và giải nghĩa các kết quả thu được. Hiểu được những ví dụ này nghiên cứu sinh có thể vận dụng để phân tích các bộ số liệu thí nghiệm thực tế và có thể sử dụng được các phần mềm thống kê khác để xử lý số liệu.

2.1. GIỚI THIỆU PHẦN MỀM THỐNG KÊ

Phần này giới thiệu sơ bộ về cách khởi động, nhập dữ liệu, khai lệnh và kết quả phân tích thống kê trên một phần mềm đại diện là SAS. Khi đã nắm được cách phân tích thống kê trên phần mềm SAS thì việc chuyển sang sử dụng các phần mềm máy tính khác có thể thực hiện một cách dễ dàng. File ví dụ và các câu lệnh của chương trình SAS có thể truy cập từ địa chỉ <https://sites.google.com/site/doduluc/download> (mục SAS data examples). Các minh họa trong giáo trình này dùng phần mềm SAS (version 8) nên có thể có giao diện khác so với giao diện trên máy tính của bạn.

2.1.1. Khởi động phần mềm

Từ menu Start của Windows chọn: Programs ⇒ The SAS System ⇒ The SAS System for Windows V8

Các cửa sổ (windows) chính của phần mềm gồm:

Editor: Cho phép tạo ra các dòng lệnh của một file mới, thay đổi và sửa chữa các file đã có sẵn. Toàn bộ số liệu được quản lý và thao tác thông qua cửa sổ này. Mọi thay đổi câu lệnh trong cửa sổ này có thể làm thay đổi cơ sở dữ liệu ban đầu hoặc/ và kết quả xử lý.

Output: Hiển thị kết quả xử lý dữ liệu thông qua các câu lệnh ở cửa sổ Editor. Kết quả xử lý có thể lưu lại trên máy tính hoặc có thể in trực tiếp ra giấy. Tuy nhiên việc in trực tiếp kết quả từ cửa sổ Editor không được khuyến cáo vì có thể gây lãng phí và khó theo dõi vì có rất nhiều khoảng trống.

Log: Hiển thị các sự kiện liên quan đến quá trình xử lý dữ liệu, bao gồm các câu lệnh thực hiện, thời gian thực hiện, các lưu ý, các cảnh báo, các thông báo về lỗi và vị trí lỗi (nếu có).

2.1.2. Nhập dữ liệu

Ví dụ sau đây được dùng để hướng dẫn cách nhập và phân tích số liệu.

Ví dụ 2.1: Khối lượng (g/con) của 16 chuột cái tại thời điểm cai sữa như sau:

54,1	49,8	24,0	46,0	44,1	34,0	52,6	54,4
56,1	52,0	51,9	54,0	58,0	39,0	32,7	58,5

Tính các tham số thống kê mô tả của đàn chuột trên bằng phần mềm SAS.

Có hai cách để nhập dữ liệu vào phần mềm SAS:

- (1) Nhập trực tiếp qua cửa sổ Editor hoặc
- (2) Nhập gián tiếp thông qua menu Import.

a. Nhập dữ liệu trực tiếp qua cửa sổ Editor

Qua cửa sổ Editor có thể lập cơ sở dữ liệu (tên cơ sở dữ liệu, tên biến, số liệu thô...) trực tiếp trên phần mềm SAS. Cách nhập này giúp người sử dụng có thể trực tiếp tạo được bộ số liệu một cách trực quan. Bên cạnh những ưu điểm thì hạn chế lớn nhất của cách nhập này là mất nhiều thời gian thao tác để nhập dữ liệu, khó kiểm soát, hiệu chỉnh dữ liệu và không sử dụng được các bộ dữ liệu có sẵn dưới dạng cơ sở dữ liệu. Cách nhập số liệu này chỉ phù hợp đối với các bộ dữ liệu bé.

Các câu lệnh để tạo bộ số liệu trực tiếp trong phần mềm SAS như sau:

```
OPTIONS PAGESIZE = 60 LINESIZE = 80; ❶
DATA SAS1;                               ❷
INPUT KL;                                ❸
CARDS;                                   ❹
54.1                                     ❺
49.8
.
32.7
58.5
;
```

Trong đó: ❶ **OPTIONS PAGESIZE** xác định số dòng tối đa in trên một trang giấy của phần kết quả (*trong ví dụ này là 60*) và **LINESIZE** xác định số ký tự tối đa trên một dòng in (*trong ví dụ này là 80*), ❷ **DATA:** Yêu cầu SAS tạo bộ số liệu trong bộ nhớ đệm của SAS và tên của bộ số liệu được tạo ra là **SAS1**, ❸ **INPUT:** Yêu cầu SAS tạo ra một biến (một cột dữ liệu) có tên là **KL**, ❹ **CARDS:** Thông báo cho SAS các số

liệu sẽ xuất hiện sau câu lệnh này. Dùng phím Enter để xuống hàng nhằm phân biệt kết thúc một số liệu, ❸ Các số liệu thô đưa vào SAS để phân tích và ❹ Kết thúc việc nhập số liệu thô bằng dấu (;).

b. Nhập dữ liệu gián tiếp qua menu IMPORT

Nhập dữ liệu gián tiếp qua menu IMPORT của chương trình SAS cho phép tiết kiệm thời gian thao tác để nhập dữ liệu, dễ kiểm soát, hiệu chỉnh dữ liệu và sử dụng được các bộ dữ liệu có sẵn dưới dạng cơ sở dữ liệu (từ excel, access...). Bạn đọc được khuyến cáo sử dụng cách nhập dữ liệu này (trong giáo trình này chỉ đề cập đến cách nhập dữ liệu này).

Để nhập dữ liệu gián tiếp thông qua menu Import cần phải có bộ dữ liệu ở dưới dạng file Excel, nếu dữ liệu đang lưu trữ trong Access thì có thể dễ dàng chuyển đổi qua Excel. Để có thể hoàn tất việc nhập dữ liệu gián tiếp thành công cần lưu ý:

- Tên của biến (tên cột) khuyến cáo không dài quá 7 ký tự, không có khoảng trống giữa các ký tự và không dùng các ký tự đặc biệt, không bắt đầu bằng số;
- Các ô không có dữ liệu không được để trống mà phải thay thế bằng dấu chấm (.);
- Trong quá trình nhập (import) dữ liệu, lỗi thao tác được thể hiện ở cửa sổ LOG (nếu có). Cửa sổ LOG sẽ thông báo vị trí và nguyên nhân mắc lỗi để bạn đọc có thể dễ dàng khắc phục.

Các bước để nhập dữ liệu gián tiếp gồm:

Bước 1: Tạo bộ số liệu trên Excel từ ví dụ nêu trên bằng cách nhập cột số liệu vào một cột với tên là KL trên Worksheet Excel. Lưu file dữ liệu dưới tên VIDU1.XLS lên thư mục SAS trên ổ đĩa D (D:\SAS2014\VIDU1.XLS).

Bước 2: Lưu file dữ liệu vừa tạo ở Bước 1 dưới định dạng text (Text (Tab delimited) (*.txt)) bằng menu SAVE AS... trong Excel dưới tên file VIDU1.TXT vào cùng thư mục. Thoát khỏi file text vừa tạo ra. Lưu ý rằng bước import sẽ không thực hiện được nếu file text vừa tạo ra vẫn còn mở.

Bước 3: Từ phần mềm SAS chọn

File ⇒ Import Data... ⇒ Tab Delimited File (.txt) ⇒ Next ⇒ VIDU1.TXT (sử dụng Browse (D:\SAS2014) để tìm file text dữ liệu cần thiết đã tạo ra ở bước 2) ⇒ Next ⇒ VIDU1 (tạo tên cơ sở dữ liệu tại ô Member:)

Bước 4: Kiểm tra các thông báo ở cửa sổ LOG để biết thêm thông tin về việc nhập số liệu. Nếu nhập số liệu thành công sẽ có thông báo như sau trong cửa sổ LOG:

16 rows created in WORD.VIDU1	from D:\SAS2014\VIDU1.txt.
NOTE: WORD.VIDU1 was successfully created	

2.1.3. Tạo câu lệnh

Để có thể tính toán được các tham số thống kê, các câu lệnh cần phải được tạo ra ở cửa sổ Editor theo các bước sau đây:

Bước 1: Kích hoạt cửa sổ EDITOR của chương trình SAS bằng cách click chuột vào cửa sổ này, ấn phím F4 trên bàn phím hoặc từ menu Run của SAS chọn Recall Last Submit (thực hiện bước này ngay sau khi Import dữ liệu, xem mục b phần 2.1.2.) để có được chương trình như sau:

```
data WORK.VIDU1          ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile  'D:\SAS2014\VIDU1.txt'  delimiter='09'x  MISSOVER  DSD  lrecl=32767
firstobs=2 ;
informat KL best32. ;
format  KL best12. ;
input
KL
;
if _ERROR_ then call symput('_EFIERR_',1); /* set ERROR detection macro
variable */
run;
```

Bước 2: Xóa hai hàng cuối cùng ở bước 1

```
if _ERROR_ then call symput('_EFIERR_',1); /* set ERROR detection macro
variable */
run;
```

Sử dụng thủ tục PROC (bắt đầu) để khai các lệnh phân tích dữ liệu và RUN. Tiếp tục phần trên ta khai các lệnh hoàn chỉnh như sau:

```
data WORK.VIDU1          ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile  'D:\SAS2014\VIDU1.txt'  delimiter='09'x  MISSOVER  DSD  lrecl=32767
firstobs=2 ;
    informat KL best32. ;
    format  KL best12. ;
input
    KL
;

```

```

PROC MEANS N MEAN STD STDERR CV;
VAR KL;
RUN;

```

Có thể lưu toàn bộ câu lệnh *File* ⇒ *Save* vào thư mục *D:\SAS2014* với tên file VIDU1.SAS để sử dụng cho các bài tập tiếp theo.

2.1.4. Kết quả phân tích

Sau khi thực hiện chương trình bằng cách chọn các menu lần lượt trên chương trình SAS (Run ⇒ Submit) hoặc chọn biểu tượng run man trên thanh công cụ, kết quả phân tích thống kê theo các lệnh đã khai báo cho ra kết quả.

Trong ví dụ trên từ cửa sổ Output của SAS có thể thấy kết quả như sau:

The MEANS Procedure				
Analysis Variable: KL				
N	Mean	Std Dev	Std Error	Coeff of Variation
①	②	③	④	⑤
16	47.5750000	10.1621848	2.5405462	21.3603464

Trong đó: ① Dung lượng mẫu ② Giá trị trung bình ③ Độ lệch chuẩn ④ Sai số chuẩn ⑤ Hệ số biến động

Xét lại Ví dụ 2.1 giả sử rằng 16 chuột này thuộc 2 giống khác nhau (A và B) và số liệu thu được như sau:

A	A	B	B	B	B	A	A
54,1	49,8	24,0	46,0	44,1	34,0	52,6	54,4
A	B	B	A	A	A	A	A
56,1	52,0	51,9	54,0	58,0	39,0	32,7	58,5

Có thể cấu trúc lại dữ liệu để thể hiện khối lượng của từng chuột thuộc giống A hoặc B. Bạn đọc cũng có thể tạo file dữ liệu với việc bổ sung thêm một cột “giống” (xem cấu trúc file VIDU1A.XLS). Sử dụng lệnh CLASS để tính các tham số thống kê mô tả đối với từng giống (A và B). Câu lệnh sử dụng như sau:

```

data WORK.VIDU1A ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */

```

```

infile 'D:\SAS2014\VIDU1A.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
    informat KL best32. ;
    informat GIONG $1. ;
    format KL best12. ;
    format GIONG $1. ;
input
    KL
    GIONG $
;
PROC MEANS N MEAN STD STDERR CV;
VAR KL;
CLASS GIONG; ❶
RUN;

```

Trong đó: ❶ CLASS dùng để tính các tham số thống kê mô tả theo từng giống “GIONG”.

Kết quả từ SAS:

The MEANS Procedure

Analysis Variable: KL

	N			Coeff of		
GIONG	Obs	N	Mean	Std Dev	Std Error	Variation
A	10	10	50.9200000	8.4607328	2.6755186	16.6157361
B	6	6	42.0000000	11.0129015	4.4959982	26.2211941

2.1.5. Chuyển kết quả từ phần mềm SAS qua một định dạng khác (Output export)

Kết quả (Output) xử lý từ phần mềm SAS có thể đổi qua một số định dạng khác như: Excel (.xls), văn bản Word (.doc), văn bản text (.txt), trình duyệt web (.html). Có hai mục đích chính của chuyển đổi định dạng kết quả từ SAS qua các định dạng khác là (1) Người sử dụng không có phần mềm SAS vẫn có thể đọc kết quả một cách dễ dàng và (2) Tóm tắt và trình bày các kết quả xử lý nhanh và chính xác nhất có thể.

a. Chuyển định dạng trực tiếp từ cửa sổ OUTPUT đang kích hoạt

File ⇒ Save as... ⇒ chọn tên file (file name:) ⇒ chọn định dạng (RTF File)

b. Chuyển định dạng thông qua câu lệnh ODS

Với câu lệnh ODS ta có thể tạo định dạng mong muốn (XLS, DOC, RTF, HTML...) và lưu file kết quả đó vào vị trí mong muốn trên máy tính, ví dụ:

```
ODS HTML FILE = "D:\SAS2014\KQVIDU1A.XLS"; ❶  
PROC MEANS N MEAN STD STDERR CV;  
VAR KL;  
CLASS GIONG;  
RUN;  
ODS HTML CLOSE; ❷
```

Trong đó: ❶ Câu lệnh ODS sẽ tạo ra file dưới định dạng excel có tên KQVIDU1A.XLS tại thư mục SAS2014 trên ổ D. Thay đổi định dạng khác bằng cách thay đổi phần mở rộng của file (Ví dụ: KETQUA.DOC nếu muốn có định dạng Word), ❷ dùng để kết thúc lệnh ODS.

Kết quả thu được trên file excel như sau:

The MEANS Procedure						
Analysis Variable: KL						
GIONG	N Obs	N	Mean	Std Dev	Std Error	Coeff of Variation
A	10	10	50.92	8.4607328	2.6755186	16.6157361
B	6	6	42	11.0129015	4.4959982	26.2211941

2.2. PHÂN TÍCH PHƯƠNG SAI (ANOVA)

2.2.1. Kiểm định điều kiện phân tích phương sai

Phân tích phương sai (Analysis of Variance - ANOVA) là công cụ hữu ích để so sánh nhiều giá trị trung bình. Phân tích phương sai được áp dụng khi biến phụ thuộc (chỉ tiêu theo dõi thí nghiệm) có dạng biến liên tục và biến độc lập (yếu tố ảnh hưởng) thuộc dạng biến phân loại. Nếu chỉ có số liệu từ hai nhóm (2 nghiệm thức) thì có thể áp dụng kiểm định T để so sánh (không giới thiệu trong giáo trình này) thay thế cho ANOVA. Tuy nhiên, nếu có hơn hai nhóm thì phải sử dụng phân tích phương sai. Điều kiện của bài toán phân tích phương sai là: (1) Các nghiệm thức có số liệu phối chuẩn và (2) Các nghiệm thức có phương sai đồng nhất. Đó là những điều kiện cần được kiểm định trước khi phân tích ANOVA.

a. Kiểm định phân phối chuẩn

Để phân tích phương sai các biến định lượng, ta phải giả thiết rằng số liệu thu thập được đều tuân theo phân phối chuẩn và các phương sai đồng nhất. Nếu số liệu không tuân theo phân phối chuẩn thì ANOVA sẽ không có hiệu lực. Do đó trước khi phân tích ANOVA ta cần kiểm tra liệu số liệu có tuân theo phân phối chuẩn hay không. Shapiro-Wilk và Kolmogorov-Smirnov là 2 phương pháp thường được sử dụng để kiểm tra phân phối chuẩn của số liệu. Shapiro-Wilk sử dụng khi dung lượng mẫu bé hơn 2000

còn Kolmogorov-Smirnov được sử dụng khi dung lượng mẫu lớn hơn 2000 (Hun Myoung Park, 2008).

Giả thiết: H_0 : Số liệu có phân phối chuẩn và H_1 : Số liệu không có phân phối chuẩn.

Ví dụ 2.2: Tăng khối lượng trung bình (g/ngày) của 36 con lợn nuôi vỗ béo giống Landrace được rút ngẫu nhiên từ một trại chăn nuôi. Số liệu thu được như sau:

577 596 594 612 600 584 618 627 588 601 606 559 615 607 608 591 565 586
621 623 598 602 581 631 570 595 603 605 616 574 578 600 596 619 636 589

Kiểm định xem số liệu trên có tuân theo phân phối chuẩn không? Số liệu được nhập vào file dưới định dạng excel với 1 cột duy nhất (KL) với tên file VIDU2.XLS; sau đó được chuyển qua định dạng text với tên file VIDU2.TXT (chi tiết xem mục 2.1.2 b).

SAS CODE lưu tại file VIDU2.SAS:

```
data WORK.VIDU2 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU2.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat KL best32. ;
  format KL best12. ;
  input
    KL
  ;
  ODS HTML FILE = 'D:\SAS2014\KQVIDU2.HTML' ; ❶
  PROC UNIVARIATE NORMAL; ❷
  VAR KL; ❸
  RUN;
  ODS HTML CLOSE;
```

Trong đó: ❶ Lưu file kết quả KQVIDU2.HTML theo định dạng HTML tại D:\SAS2014, ❷ PROC UNIVARIATE NORMAL câu lệnh kiểm tra phân bố chuẩn (NORMAL) của ❸ cột số liệu khối lượng (VAR KL).

Kết quả từ SAS KQVIDU2.HTML:

The UNIVARIATE Procedure			
Variable: KL			
Moments			
N	36	Sum Weights	36
Mean	599.194444	Sum Observations	21571
Std Deviation	18.6560131	Variance	348.046825
Skewness	-0.1258564	Kurtosis	-0.4077877
Uncorrected SS	12937405	Corrected SS	12181.6389

Moments			
Coeff Variation	3.1135157	Std Error Mean	3.10933552

Basic Statistical Measures ❶			
Location ❷		Variability ❸	
Mean	599.1944	Std Deviation	18.65601
Median	600.0000	Variance	348.04683
Mode	596.0000	Range	77.00000
		Interquartile Range	26.50000

Trong đó: **❶** Tham số thống kê mô tả **❷** chỉ vị trí và **❸** mức độ biến động của số liệu

Tests for Location: $\mu_0=0$				
Test	Statistic		p Value	
Student's t	t	192.7082	Pr > t	<.0001
Sign	M	18	Pr >= M	<.0001
Signed Rank	S	333	Pr >= S	<.0001

Tests for Normality ❶				
Test	Statistic		p Value	
Shapiro-Wilk ❷	W	0.99129	Pr < W	0.9918
Kolmogorov-Smirnov ❸	D	0.057007	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.01366	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.094344	Pr > A-Sq	>0.2500

Trong đó: **❶** Kiểm tra phân bố chuẩn của cột số liệu; **❷** Khi sử dụng phép thử Shapiro-Wilk ($n = 36 < 2000$), giá trị $P = 0,9918 > 0,05$, như vậy H_0 được chấp nhận. Kết luận: Số liệu tuân theo phân phối chuẩn; **❸** Nếu $n > 2000$, sử dụng phép thử Kolmogorov-Smirnov.

Quantiles (Definition 5)

Quantile	Estimate
100% Max	636.0
99%	636.0
95%	631.0
90%	623.0
75% Q3	613.5
50% Median	600.0
25% Q1	587.0
10%	574.0
5%	565.0
1%	559.0
0% Min	559.0

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
559	12	621	19
565	17	623	20
570	25	627	8
574	30	631	24
577	1	636	35

b. Kiểm định sự đồng nhất của phương sai

Trong thực tế, việc kiểm định phương sai đồng nhất được thực hiện đồng thời với kiểm tra phân bố chuẩn khi tiến hành phân tích phương sai. Đối với ANOVA, việc kiểm định phân bố chuẩn và phương sai đồng nhất được kiểm tra thông qua sự phân bố của phân dư (ε) của mô hình phân tích. Bạn đọc tham khảo *phần a, mục 2.2.3* để biết thêm chi tiết các câu lệnh sử dụng để kiểm định phân dư (ε).

2.2.2. So sánh cặp đôi các giá trị trung bình

Khi kết quả phân tích ANOVA cho thấy có sự sai khác có ý nghĩa thống kê ($P < 0,05$) giữa các nghiệm thức thì tiến hành so sánh cặp đôi (pairwise comparison) để tìm sự khác nhau có ý nghĩa giữa các giá trị trung bình cụ thể nào. Với kiểm định 2 phía ta có giả thiết $H_0: \mu_1 = \mu_2 = \dots = \mu_a$ và đối thiết $H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_a$ (μ là trung bình của quần thể ở công thức thứ 1, 2,..., a). Khi kết quả phân tích ANOVA cho thấy $P < 0,05$,

có nghĩa là H_0 bị bác bỏ vì có ít nhất hai giá trị trung bình sai khác nhau. Lúc đó đối thiết H_1 được chấp nhận và so sánh cặp đôi sẽ cho phép tìm được sự khác nhau cụ thể giữa các giá trị trung bình của các nghiệm thức 1, 2,..., a.

Có nhiều phương pháp so sánh cặp đôi khác nhau, trong đó một số phương pháp so sánh cặp đôi thường được sử dụng như kiểm định Student-Newman-Keuls, LSD (Least Significant Difference), Tukey, Scheffe, kiểm định đa phạm vi Duncan. Việc lựa chọn kiểm định nào tùy thuộc vào đặc điểm của bộ dữ liệu và yêu cầu về tính chắc chắn của sự sai khác.

Xét một thí nghiệm có một yếu tố thí nghiệm với 4 công thức thí nghiệm ($a = 4$), sẽ có tổng số 6 lần so sánh cặp. Nếu mỗi lần so sánh một cặp với sai số $P = 0,05$; sai số của toàn bộ 6 phép thử sẽ có sai số là $1 - (1 - 0,05)^6 = 0,24$.

Với thí nghiệm có a nghiệm thức và mỗi phép so sánh cặp với sai số là α thì $a * \alpha$ là mức ý nghĩa tổng số (overall significance level) hoặc sai số của thí nghiệm (experimental wise error rate) hoặc sai số của toàn bộ các phép thử (family error rate). Sai số của toàn bộ phép thử được tính toán như thế nào?

Ví dụ: Nếu sai số của một cặp so sánh là $\alpha = 0,05$ và có $a = 4$ ta sẽ có $m = 6$ cặp so sánh. Như vậy sai số của 6 phép thử sẽ là $1 - (1 - \alpha)^m = 1 - (1 - 0,05)^6 = 0,264908$. Nghĩa là xác suất bác bỏ giả thiết H_0 khi nó đúng là 26,4908% (sai lầm loại I).

Sau đây là một số lưu ý khi dùng các phương pháp so sánh cặp đôi khác nhau:

Kiểm định LSD – Chỉ sử dụng để so sánh giữa các cặp đã được thiết kế trước khi thực hiện thí nghiệm. Không phù hợp để:

(1) Kiểm định LSD không phù hợp để so sánh nhiều công thức thí nghiệm vì khi có nhiều nghiệm thức thì số cặp để so sánh sẽ tăng lên rất nhanh (5 nghiệm thức có 10 cặp để so sánh, 10 nghiệm thức có 45 cặp so sánh, 15 nghiệm thức có 105 cặp).

(2) Kiểm định LSD không phù hợp để so sánh số cặp nhiều hơn số bậc tự do của công thức thí nghiệm ($a - 1$). Ví dụ thí nghiệm có 4 công thức ($a = 4$) thì số cặp có thể so sánh tối đa là $4 - 1 = 3$.

(3) Nếu muốn kết luận có sự sai khác giữa 2 giá trị trung bình (μ_i và μ_j) thì $\mu_i - \mu_j > LSD$, trong đó $LSD = t_{\alpha/2, dfe} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$. LSD được xác định dựa vào giá trị t ở mức $\alpha/2$, bậc tự do của sai số ngẫu nhiên (dfe) và dung lượng mẫu (n) ở công thức thứ i và j.

(4) LSD bị chỉ trích vì sự kiểm soát không hiệu quả đối với sai lầm loại I.

Kiểm định Tukey – Được sử dụng để so sánh cặp. Sử dụng Tukey-HSD (Honest Significant Difference) trong trường hợp dung lượng mẫu bằng nhau và Tukey-Kramer trong trường hợp dung lượng mẫu không bằng nhau. Kiểm định Tukey-HSD được nhiều phần mềm xử lý số liệu áp dụng cho cả trường hợp dung lượng mẫu không bằng nhau vì

vậy nhiều tài liệu không còn phân biệt giữa Tukey-HSD và Tukey-Kramer. Trong phần mềm SAS cũng chỉ có option Tukey, không có lựa chọn cho Tukey-HSD hay Tukey-Kramer. Phép kiểm định Tukey phù hợp trong trường hợp chỉ so sánh giữa các cặp với nhau.

- Giá trị tới hạn đối với kiểm định Tukey-HSD trong trường hợp dung lượng mẫu bằng nhau: $CR = q_{\alpha, p, dfe} \sqrt{MSE / n}$;

- Giá trị tới hạn của kiểm định Tukey-Kramer trong trường hợp dung lượng mẫu không bằng nhau: $CR = q_{\alpha, dfa, dfe} \sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$.

Trong đó p là giá trị trong bảng Tukey tương ứng ở mức α , bậc tự do của nghiệm thức (dfa), bậc tự do sai số ngẫu nhiên (dfe), sai số ngẫu nhiên (MSE) và dung lượng mẫu (n) ở nghiệm thức i và j .

Giá trị tới hạn trong kiểm định Tukey với trường hợp dung lượng mẫu bằng nhau luôn là một giá trị cố định; ngược lại khi dung lượng mẫu không bằng nhau giá trị CR sẽ thay đổi theo từng cặp so sánh.

Tukey là phép thử dùng để so sánh tất cả các cặp có thể trong thí nghiệm, hay được sử dụng vì sai lầm loại I luôn được kiểm soát ($\alpha_{FWE} < 0,05$) và độ mạnh của phép thử ($1 - \beta$) cao hơn so với các phép thử trong tình huống tương tự. Mặc dù vậy, đối với các thí nghiệm mà không cần so sánh tất cả các cặp có thể trong thí nghiệm thì có thể sử dụng các phép thử khác như Dunnett hoặc Bonferroni.

Kiểm định Dunnett – Là phép thử có cách tiếp cận tương tự như đối với trường hợp Tukey. Tuy nhiên phép thử này được sử dụng trong trường hợp so sánh các công thức thí nghiệm với đối chứng được so sánh.

Kiểm định Bonferroni – Là phương pháp so sánh đơn giản và kiểm soát tốt sai lầm loại I. Nhược điểm của phương pháp này là tính bảo thủ và sai lầm loại II lớn. Bonferroni luôn kiểm soát được sai số của thí nghiệm ở mức α_{FWE} (thường $\alpha = 0,05$). Sai số của từng phép so sánh cặp sẽ là $\alpha_B = \alpha_{FWE}/m$ Nếu số cặp so sánh (m) lớn thì α_B sẽ rất bé. Sự chặt chẽ của phương pháp này chính là điểm yếu vì khi α_B bé (nếu m tăng lên) thì cơ hội tìm thấy sự sai khác giữa 2 nghiệm thức sẽ giảm xuống mặc dù có sự sai khác thật sự giữa 2 nghiệm thức.

Kiểm định Duncan – Phù hợp để so sánh tất cả các cặp có thể. Nếu đối với phương pháp LSD chỉ có một giá trị LSD duy nhất để so sánh tất cả các cặp thì Duncan lại có từng giá trị DMRT (Duncan's Multiple Range Test) cụ thể cho từng cặp nhất định. $R_p = r_{\alpha, p, v} \sqrt{MSE / n}$; trong đó r là giá trị trong bảng Duncan tương ứng ở mức α , khoảng cách giữa 2 giá trị trung bình (p), bậc tự do sai số ngẫu nhiên (dfe), sai số ngẫu nhiên (MSE) và dung lượng mẫu trong một nghiệm thức (n).

Kiểm định Scheffe: Có 3 bước được áp dụng để so sánh cặp đôi theo phương pháp của Scheffe gồm:

- (1) Xác định giá trị (tuyệt đối) sai khác giữa các nghiệm thức d
- (2) Xác định giá trị Scheffe (CR)

$$CR = \sqrt{(a-1)F_{\alpha, dfa, dfe}MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

Trong đó, a : số công thức thí nghiệm, giá trị F ở α : mức độ sai khác (thường được chọn 0,05), d_{fa} và d_{fe} là bậc tự do của nghiệm thức $(a - 1)$ và bậc tự do của sai số ngẫu nhiên $(n_1 + n_2 + \dots + n_a - a)$, MSE : trung bình bình phương của sai số ngẫu nhiên, n_i và n_j là dung lượng mẫu tương ứng ở công thức i và j .

(3) Thể hiện sự sai khác giữa các nghiệm thức. Nếu giá trị chênh lệch giữa 2 nghiệm thức (d) bé hơn giá trị lý thuyết (CR) ta sẽ có kết luận không có sự sai khác giữa 2 nghiệm thức này

Phép kiểm định Tukey được sử dụng rộng rãi hơn Scheffe. Sử dụng kiểm định Tukey, xác suất bác bỏ H_0 cao hơn nhưng xác suất mắc sai lầm loại I cũng cao hơn. Với Scheffe, xác suất bác bỏ giả thuyết H_0 thấp hơn nhưng xác suất mắc sai lầm cũng thấp hơn.

Nếu kiểm định Tukey chỉ có thể so sánh các cặp với nhau thì kiểm định Scheffe, ngoài so sánh các cặp với nhau còn có thể dùng để so sánh các tương phân (contrast), tức là các tổ hợp tuyến tính của các nghiệm thức.

Phép kiểm định nào phù hợp để so sánh các nghiệm thức trong một thí nghiệm? Bạn đọc có thể tham khảo một số gợi ý sau:

- So sánh tất cả các cặp với nhau (pairwise)
 - + Dung lượng mẫu bằng nhau giữa các nghiệm thức: kiểm định Tukey
 - + Dung lượng mẫu không bằng nhau giữa các nghiệm thức: kiểm định Tukey-Kramer or Scheffé
- So sánh một số công thức với nhau (not pairwise)
 - + Với nhóm đối chứng: kiểm định Dunnett
 - + Với một số công thức dự kiến trước: kiểm định Bonferroni
 - + Với một số công thức không dự kiến trước: kiểm định Scheffé

Các trường hợp so sánh cặp đôi có thể thấy được trong các ví dụ phân tích ANOVA với các mô hình khác nhau ở trong mục 2.2.3.

2.2.3. Các mô hình phân tích phương sai và so sánh cặp đôi

a. Thí nghiệm một nhân tố hoàn toàn ngẫu nhiên

Xét trường hợp đơn giản nhất đối với bài toán phân tích phương sai là *thí nghiệm một yếu tố hoàn toàn ngẫu nhiên (Completely Randomized Design - CRD)*. Chỉ có một yếu tố duy nhất trong thí nghiệm là biến độc lập, các yếu tố phi thí nghiệm còn lại được coi là có tác động như nhau đến đối tượng thí nghiệm.

Ví dụ 2.3: Theo dõi tăng khối lượng của 20 con cá (g) trong một thí nghiệm với 5 công thức nuôi (A, B, C, D và E). Hãy cho biết tăng khối lượng của cá ở các công thức nuôi. Nếu có sự khác nhau, tiến hành so sánh sự sai khác của từng cặp giá trị trung bình.

A	B	C	D	E
950	430	700	1.000	900
850	450	900	900	1.000
850	400	750	900	950
900	420	700	900	950

Mô hình phân tích:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Trong đó: y_{ij} : quan sát thứ j ở công thức i,

μ : trung bình chung,

α_i : ảnh hưởng của công thức i

ε_{ij} : sai số ngẫu nhiên; các ε_{ij} độc lập, phân phối chuẩn $N(0, \sigma^2)$.

Như đã đề cập ở *phần b mục 2.2.1*, việc kiểm định phân bố chuẩn và phương sai đồng nhất được kiểm tra thông qua sự phân bố của phân dư (ε) của mô hình phân tích. Tiến hành kiểm định phân dư (ε) của mô hình phân tích bằng các câu lệnh trong SAS như sau:

SAS CODE (số liệu được nhập vào file VIDU3.XLS sau đó chuyển qua file VIDU3.TXT với cột khẩu phần KP và cột khối lượng KL):

```
data WORK.VIDU3 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU3.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat KP $1. ;
  informat KL best32. ;
  format KP $1. ;
  format KL best12. ;
  input
    KP $
```

```

KL
;
PROC GLM; ❶
  CLASS KP;
  MODEL KL = KP;

  OUTPUT OUT = DLMOI R = PHANDU; ❷
RUN;
ODS HTML FILE = 'D:\SAS2014\PBCVIDU2.HTML' ;
PROC UNIVARIATE NORMAL DATA = DLMOI; ❸
  VAR PHANDU; ❹
RUN;
ODS HTML CLOSE;

```

Trong đó: ❶ Thủ tục (procedure) GLM (General Linear Model) sẽ được đề cập chi tiết ở *phần 2.2.3 mục c*; ❷ OUTPUT OUT Câu lệnh tạo ra bộ số liệu mới có tên DLMOI từ bộ số liệu ban đầu thêm một cột số liệu mới PHANDU, giá trị trong cột số liệu PHANDU này chính là các giá trị phân dư R (ε_{ij}) trong mô hình nêu trên; ❸ UNIVARIATE NORMAL câu lệnh kiểm tra phân bố chuẩn (NORMAL) của ❹ cột phân dư (VAR PHANDU).

Phản kiểm định có kết quả như sau:

The UNIVARIATE Procedure
Variable: PHANDU

Tests for Normality ❶				
Test	Statistic		p Value	
Shapiro-Wilk ❷	W	0.904978	Pr < W	0.0512
Kolmogorov-Smirnov	D	0.160193	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.108097	Pr > W-Sq	0.0845
Anderson-Darling	A-Sq	0.637719	Pr > A-Sq	0.0859

Trong đó: ❶ Kiểm tra phân bố chuẩn của số liệu và sự đồng nhất của phương sai thông qua kiểm định giá trị phân dư (ε_{ij}); ❷ Sử dụng phép thử Shapiro-Wilk ($n < 2.000$), giá trị $P = 0,0512 > 0,05$, như vậy H_0 được chấp nhận. Kết luận: Số liệu tuân theo phân phối chuẩn và phương sai đồng nhất, điều kiện của bài toán phân tích phương sai được thỏa mãn.

Phân tích phương sai (ANOVA) để tìm sự sai khác (nếu có) giữa các nghiệm thức. Nếu có sự sai khác giữa các nghiệm thức thì tiến hành so sánh cặp (xem chi tiết mục 2.2.2. *So sánh cặp đôi các giá trị trung bình*).

SAS CODE:

```
data WORK.VIDU3 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU3.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat KP $1. ;
  informat KL best32. ;
  format KP $1. ;
  format KL best12. ;
  input
      KP $
      KL
  ;
PROC ANOVA; ❶
CLASS KP; ❷
MODEL KL = KP; ❸
MEANS KP/ DUNCAN; ❹
RUN; ❺
```

Trong đó: ❶ **PROC ANOVA** Thủ tục phân tích phương sai; ❷ **CLASS** Yếu tố thí nghiệm KP; ❸ **MODEL** Mô hình phân tích (khẩu phần-biến độc lập ảnh hưởng đến khối lượng-biến phụ thuộc); ❹ So sánh sự sai khác giữa các giá trị trung bình giữa các khẩu phần (KP) bằng phép thử DUNCAN; ❺ Kết thúc thủ tục phân tích phương sai.

Kết quả phân tích từ SAS:

The ANOVA Procedure ❶	
Class Level Information ❷	
Class ❸	Values ❺
KP	A B C D E
Levels ❹	
5	
Number of observations ❻	
20	

Trong đó: ❶ Kết quả phân tích phương sai; ❷ Thông tin về yếu tố thí nghiệm; ❸ Yếu tố thí nghiệm **KP**; ❹ Số mức trong yếu tố thí nghiệm 5; ❺ Các mức gồm A B C D E.

Dependent Variable: KL ❶					
Source ❷	DF ❸	Sum of Squares ❹	Mean Square ❺	F Value ❻	Pr > F ❽
Model ❸	4	749250.0000	187312.5000	59.09	<.0001
Error ❹	15	47550.0000	3170.0000		
Corrected Total ❺	19	796800.0000			

Trong đó: ❶ Bảng phân tích phương sai của Biến phụ thuộc KL; ❷ Nguồn biến động; ❸ Mô hình phân tích; ❹ Sai số ngẫu nhiên; ❺ Tổng số; ❻ Bậc tự do; ❼ Tổng bình phương; ❽ Trung bình bình phương; ❾ Giá trị F; ❿ Xác suất.

R-Square ❶	Coeff Var ❷	Root MSE ❸	KL Mean ❹
0.940324	7.126931	56.30275	790.0000

Trong đó: ❶ Hệ số xác định R²; ❷ Hệ số biến động Cv của toàn bộ quan sát; ❸ Căn bậc 2 của trung bình bình phương (SE); ❹ Trung bình cộng của toàn bộ quan sát.

Source ❶	DF	Anova SS	Mean Square	F Value	Pr > F ❷
KP	4	749250.0000	187312.5000	59.09	<.0001

Trong đó: ❶ Biến động với từng yếu tố thí nghiệm. Trong trường hợp thí nghiệm có 1 yếu tố duy nhất, các giá trị của yếu tố thí nghiệm chính bằng của mô hình (xem chi tiết *Bảng phân tích phương sai*); ❷ Giá trị P <0,0001, H₀ bị bác bỏ và H₁ được chấp nhận. Kết luận: tăng khối lượng của cá có sự sai khác giữa các nghiệm thức.

Vì bác bỏ giả thiết H₀ và chấp nhận đối thiết H₁, bước tiếp theo là so sánh cặp đôi. Trong ví dụ này so sánh cặp đôi được thực hiện theo phương pháp Duncan. Bạn đọc có thể sử dụng các phương pháp so sánh cặp đôi khác nhau phù hợp với từng thí nghiệm (mục 2.2.2):

Duncan's Multiple Range Test for KL ❶	
NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.	
Alpha ❷	0.05
Error Degrees of Freedom	15
Error Mean Square	3170

Number of Means ③	2	3	4	5
Critical Range ④	84.86	88.95	91.50	93.23

Means with the same letter are not significantly different. ⑤			
Duncan Grouping ⑥	Mean	N	KP
A	950.00	4	E
A			
A	925.00	4	D
A			
A	887.50	4	A
B	762.50	4	C
C	425.00	4	B

Trong đó: **①** So sánh cặp bằng phương pháp Duncan đối với chỉ tiêu khối lượng *KL*; **②** mức α ; **③** Số công thức; **④** Khoảng ý nghĩa; **⑤** Sự sai khác giữa các nghiệm thức; **⑥** Nhóm Duncan.

Kết quả có thể trình bày như sau:

Nghiệm thức	A	B	C	D	E
Trung bình	887,50 ^a	425,00 ^c	762,50 ^b	925,00 ^a	950,00 ^a

Những giá trị trung bình có chung chữ cái (theo so sánh Duncan) thì sai khác nhau không có ý nghĩa thống kê và ngược lại những giá trị trung bình không có chung chữ cái thì sai khác nhau có ý nghĩa thống kê ($P < \alpha$).

Chi tiết về so sánh cặp và trình bày kết quả, bạn đọc có thể tham khảo trang 57 chương 4 Giáo trình Thiết kế thí nghiệm của Nguyễn Đình Hiền và Đỗ Đức Lực (2007).

b. Thí nghiệm một yếu tố khối ngẫu nhiên đầy đủ

Xem xét một thí nghiệm mà đối tượng thí nghiệm chịu tác động đồng thời của một yếu tố chính và một yếu tố phụ là khối theo mô hình *khối ngẫu nhiên đầy đủ* (Completely Randomized Block Design - CRBD).

- Thí nghiệm khối ngẫu nhiên đầy đủ với một lần lặp lại

Ví dụ 2.4: Nghiên cứu số lượng tế bào lymphô ở chuột ($\times 1000$ tế bào/mm³ máu) được sử dụng 4 loại thuốc khác nhau (A, B, C và D; thuốc D là placebo) qua 5 lứa; số liệu thu được trình bày ở bảng dưới. Cho biết ảnh hưởng của thuốc đến tế bào lymphô?

Thuốc	Lứa 1	Lứa 2	Lứa 3	Lứa 4	Lứa 5
A	7,1	6,1	6,9	5,6	6,4
B	6,7	5,1	5,9	5,1	5,8
C	7,1	5,8	6,2	5,0	6,2
D	6,7	5,4	5,7	5,2	5,3

Mô hình phân tích:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$

Trong đó: y_{ij} : quan sát thứ i của yếu tố ở khối thứ j ,

μ : trung bình chung,

α_i : ảnh hưởng của mức i của yếu tố,

β_j : ảnh hưởng của khối j ,

ε_{ij} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Cấu trúc số liệu: Số liệu có thể nhập vào file VIDU4.XLS định dạng Excel sau đó chuyển file VIDU4.TXT định dạng txt với 3 cột lần lượt là thuốc (THUOC), lứa (LUA) và số lượng tế bào (TEBAO).

SAS CODE của VIDU4.SAS:

```
data WORK.VIDU4
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU4.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat THUOC $1. ;
informat LUA best32. ;
informat TEBAO best32. ;
format THUOC $1. ;
format LUA best12. ;
format TEBAO best12. ;
input
    THUOC $
    LUA
    TEBAO
;
PROC ANOVA;
CLASS THUOC LUA; ❶
```

```
MODEL TEBAO = THUOC LUA; ❷
MEANS THUOC / TUKEY; ❸
RUN;
```

Trong đó: ❶ Yếu tố thí nghiệm: 2 (THUOC LUA); ❷ Mô hình phân tích có 2 yếu tố thí nghiệm; ❸ So sánh sự sai khác giữa các thuốc bằng phương pháp TUKEY.

Kết quả từ SAS:

The ANOVA Procedure ❶					
Dependent Variable: TEBAO					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	8.24850000	1.17835714	22.20	<.0001
Error	12	0.63700000	0.05308333		
Corrected Total	19	8.88550000			

R-Square	Coeff Var	Root MSE	TEBAO Mean
0.928310	3.862501	0.230398	5.965000

Trong đó: ❶ Bảng phân tích phương sai (tham khảo ví dụ 2.3).

Source ❶	DF	Anova SS	Mean Square	F Value	Pr > F ❷
THUOC	3	1.84550000	0.61516667	11.59	0.0007
LUA	4	6.40300000	1.60075000	30.16	<.0001

Trong đó: ❶ Yếu tố thí nghiệm 2, bao gồm thuốc (THUOC) và lúa (LUA); ❷ Xác suất đối với từng yếu tố thí nghiệm. Với yếu tố Thuốc $P = 0,0007$, ta bác bỏ giả thiết H_0 và chấp nhận đối thiết H_1 . Kết luận: Thuốc có ảnh hưởng khác nhau lên tế bào lymphô của chuột.

The ANOVA Procedure	
Tukey's Studentized Range (HSD) Test for TEBAO ❶	
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.	
Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	0.053083
Critical Value of Studentized Range	4.19852

Minimum Significant Difference	0.4326
--------------------------------	--------

Trong đó: ❶ So sánh cặp đôi với chỉ tiêu tế bào (TEBAO) bằng phương pháp Tukey.

Means with the same letter are not significantly different. ❶				
Tukey Grouping		Mean	N	THUOC
	A	6.4200	5	A
	A			
B	A	6.0600	5	C
B				
B		5.7200	5	B
B				
B		5.6600	5	D

Trong đó: ❶ Kết quả so sánh cặp đôi giữa các thuốc bằng phương pháp Tukey.

Kết quả có thể trình bày như sau:

Thuốc	A	B	C	D
Trung bình	6,42 ^a	5,72 ^b	6,06 ^{ab}	5,66 ^b

Lưu ý rằng, mô hình thiết kế thí nghiệm khối ngẫu nhiên đầy đủ với một lần lặp lại sẽ không nghiên cứu được mỗi tương tác giữa 2 yếu tố thí nghiệm. Vì trong trường hợp này phần sai số ngẫu nhiên chính là phần tương tác giữa 2 yếu tố nếu đưa vào mô hình. Để có thể nghiên cứu được mỗi tương tác giữa 2 yếu tố, bạn đọc có thể tìm hiểu mô hình thí nghiệm khối ngẫu nhiên hoàn toàn với nhiều lần lặp lại dưới đây.

- *Thí nghiệm khối ngẫu nhiên hoàn toàn với nhiều lần lặp lại*

Ví dụ 2.5: Một thí nghiệm được tiến hành để xác định ảnh hưởng của 3 công thức thức ăn (A1, A2 và A3) đến tăng khối lượng trung bình trên ngày (g/ngày) của bê đực. Bê đực được cân và chia thành 4 khối dựa theo khối lượng bắt đầu thí nghiệm. Trong mỗi khối có 6 động vật thí nghiệm được chọn ra và được phân ngẫu nhiên về các nghiệm thức. Số liệu thu thập sau khi kết thúc thí nghiệm như sau:

Công thức	Khối			
	I	II	III	IV
A1	826	864	795	850
	806	834	810	845

A2	827	871	729	860
	800	881	709	840
A3	753	801	736	820
	773	821	740	835

Nếu trong một công thức - một khối có nhiều quan sát thì ngoài việc đánh giá mức độ ảnh hưởng của từng yếu tố riêng biệt ta còn có thể xác định mối tương tác theo mô hình phân tích sau:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \times \beta)_{ij} + \varepsilon_{ijk}$$

Trong đó:

y_{ijk} : quan sát thứ k của khối thứ j và nghiệm thức thứ i,

μ : trung bình chung,

α_i : ảnh hưởng của công thức thức ăn i,

β_j : ảnh hưởng của khối lượng ban đầu của khối j,

$(\alpha \times \beta)_{ij}$: tương tác giữa công thức thức ăn và khối,

ε_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$,

Cấu trúc số liệu tương tự như ví dụ 2.4: Số liệu có thể nhập vào file VIDU5.XLS định dạng excel sau đó chuyển file VIDU5.TXT định dạng.txt với 3 cột lần lượt là công thức thức ăn (CTTA), nhóm khối lượng (KHOI) và tăng khối lượng (KL).

SAS CODE của VIDU5.SAS:

```
data WORK.VIDU5
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU5.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat CTTA $2. ;
informat KHOI $4. ;
informat KL best32. ;
format CTTA $2. ;
format KHOI $4. ;
format KL best12. ;
input
CTTA $
KHOI $
KL
;
```

```

PROC ANOVA; ❶

CLASS CTTA KHOI; ❷

MODEL KL = CTTA KHOI CTTA*KHOI; ❸

RUN;

```

Trong đó: ❶ Thủ tục phân tích phương sai (ANOVA) được sử dụng để phân tích; ❷ Ảnh hưởng của 2 yếu tố thí nghiệm (CTTA KHOI); ❸ Mô hình phân tích ảnh hưởng của công thức thức ăn (CTTA) khối (KHOI) và tương tác giữa 2 yếu tố này (CTTA*KHOI).

Kết quả từ SAS:

The ANOVA Procedure ❶

Class Level Information ❷		
Class	Levels	Values
CTTA	3	A1 A2 A3
KHOI	4	I II III IV

Number of observations	24
------------------------	----

The ANOVA Procedure

Dependent Variable: KL ❸

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	49929.83333	4539.07576	25.81	<.0001
Error	12	2110.00000	175.83333		
Corrected Total	23	52039.83333			

R-Square	Coeff Var	Root MSE	KL Mean
0.959454	1.638244	13.26022	809.4167

Source ❹	DF	Anova SS	Mean Square	F Value	Pr > F ❺
CTTA	2	8025.58333	4012.79167	22.82	<.0001
KHOI	3	33816.83333	11272.27778	64.11	<.0001

Source ^④	DF	Anova SS	Mean Square	F Value	Pr > F ^⑤
CTTA*KHOI	6	8087.41667	1347.90278	7.67	0.0015

Trong đó: ① Kết quả phân tích phương sai; ② Thông tin về yếu tố thí nghiệm và các mức của từng yếu tố; ③ Kết quả phân tích đối với biến khối lượng (KL); ④ Bảng phân tích phương sai với các yếu tố ảnh hưởng công thức thức ăn (CTTA), khối (KHOI) và tương tác giữa 2 yếu tố (CTTA*KHOI); ⑤ Xác suất với yếu tố *Thức ăn* $P < 0,0001$ và tương tác (CT*KHOI) $P = 0,0015$, ta bác bỏ giả thiết H_0 và chấp nhận đối thiết H_1 .

Kết luận: Công thức ăn có ảnh đến tăng khối lượng của bê và có tương tác giữa công thức thức ăn và khối lượng bê vỗ béo ($P < 0,05$).

c. Thí nghiệm hoán vị (cross over)

Trong thiết kế thí nghiệm kiểu hoán vị, có 2 hay nhiều công thức thí nghiệm được thực hiện trên cùng một động vật thí nghiệm nhưng ở các giai đoạn khác nhau. Số liệu được thu thập trên đối tượng thí nghiệm nhiều lần tương ứng với các công thức thí nghiệm khác nhau. Việc bố trí các nghiệm thức trên một động vật thí nghiệm là ngẫu nhiên và từng động vật được xem như một khối.

Giai đoạn	Động vật thí nghiệm						
	1	2	3	4	5	...	n
1	CT2	CT1	CT2	CT3	CT1	...	CT3
2	CT1	CT3	CT3	CT2	CT3	...	CT2
3	CT3	CT2	CT1	CT1	CT2	...	CT1

Mô hình ô vuông la tinh cũng có thể được sử dụng trong trường hợp này. Đặc biệt các thí nghiệm được thiết kế trên đại gia súc. Sơ đồ dưới đây minh họa sơ đồ thiết kế thí nghiệm có 2 ô vuông la tinh được thiết kế đồng thời.

Giai đoạn	Động vật thí nghiệm						
	Ô vuông la tinh thứ 1				Ô vuông la tinh thứ 2		
	1	2	3		4	5	6
1	CT2	CT1	CT2		CT3	CT1	CT3
2	CT1	CT3	CT3		CT2	CT3	CT2
3	CT3	CT2	CT1		CT1	CT2	CT1

Trường hợp đặc biệt có 2 công thức thí nghiệm sẽ có một nhóm động vật tham gia thí nghiệm với công thức thí nghiệm thứ nhất, nhóm còn lại sẽ tham gia công thức 2. Sau một thời gian các công thức được thay đổi ngược lại.

Ví dụ 2.6: Một thí nghiệm được tiến hành nhằm nghiên cứu ảnh hưởng của 2 khẩu phần thức ăn đến sản lượng sữa. Tổng số 10 bò ở tháng tiết sữa thứ 3 và thứ 4 được chọn làm thí nghiệm. Bò được chia ngẫu nhiên thành 2 nhóm (nhóm 1 và nhóm 2).

Từng nhóm bò được thử nghiệm trên 2 công thức theo từng giai đoạn khác nhau. Đối với nhóm 1, áp dụng khẩu phần 1 ở giai đoạn 1 và khẩu phần 2 ở giai đoạn 2; đối với nhóm 2 áp dụng ngược lại. Sản lượng sữa trung bình (kg) của từng bò trong từng giai đoạn được trình bày ở bảng sau:

Nhóm 1						
Giai đoạn	Khẩu phần	Bò 1	Bò 4	Bò 5	Bò 9	Bò 10
1	1	31	34	43	28	25
2	2	27	25	38	20	19
Nhóm 2						
Giai đoạn	Khẩu phần	Bò 2	Bò 3	Bò 6	Bò 7	Bò 8
1	2	22	40	40	33	18
2	1	21	39	41	34	20

Số liệu có thể nhập vào file VIDU6.XLS định dạng excel sau đó chuyển file VIDU6.TXT định dạng txt với 5 cột lần lượt là giai đoạn (GD), khẩu phần (KP), nhóm bò (NHOM), bò (BO) và sản lượng sữa (SLS).

Nếu không đề cập đến ảnh hưởng của *nhóm bò* ta có mô hình phân tích:

$$y_{ijk} = \mu + kp_i + bo_j + e_{ijk}$$

Trong đó: y_{ijk} : quan sát ở khẩu phần thứ i, nhóm bò thứ j và của bò k

μ : trung bình chung,

kp_i : chênh lệch do ảnh hưởng của khẩu phần i,

bo_j : ảnh hưởng của bò j,

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

SAS CODE của VIDU6:

```
data WORK.VIDU6                                ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU6.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat GD best32. ;
informat KP best32. ;
informat NHOM best32. ;
informat BO best32. ;
informat SLS best32. ;
format GD best12. ;
format KP best12. ;
format NHOM best12. ;
format BO best12. ;
format SLS best12. ;
```

```

input
    GD
    KP
    NHOM
    BO
    SLS
;
PROC ANOVA; ❶
CLASS KP BO; ❷
MODEL SLS = KP BO; ❸
RUN;

```

Trong đó: ❶ Thủ tục ANOVA được sử dụng để phân tích phương sai; ❷ Yếu tố thí nghiệm gồm khẩu phần (KP), và bò (BO). Bò được coi như một khối đối với mô hình phân tích; ❸ Mô hình (MODEL) phân tích ảnh hưởng của các yếu tố thí nghiệm (KP BO) đến sản lượng sữa bò (SLS).

Kết quả từ SAS:

The ANOVA Procedure ❶

Dependent Variable: SLS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	1292.600000	129.260000	20.34	<.0001
Error	9	57.200000	6.355556		
Corrected Total	19	1349.800000			

R-Square	Coeff Var	Root MSE	SLS Mean
0.957623	8.431514	2.521023	29.90000

Source	DF	Anova SS	Mean Square	F Value	Pr > F ❷
BO	9	1234.800000	137.200000	21.59	<.0001
KP	1	57.800000	57.800000	9.09	0.0146

Trong đó: ❶ Kết quả phân tích phương sai với thủ tục ANOVA; việc đọc kết quả tương tự như ở ví dụ trên; ❷ Xác suất của từng yếu tố thí nghiệm. Trong thí nghiệm này, khẩu phần (KP) là yếu tố quan tâm chính; nghĩa là sản lượng sữa bò có sự khác biệt

giữa 2 công thức hay không? Xác suất của phép phân tích đối với công thức thức ăn $P = 0,0146 < 0,05$. Kết luận: Công thức thức ăn có ảnh hưởng đến sản lượng sữa của bò.

Nếu đề cập đến ảnh hưởng của nhóm bò ta có mô hình phân tích:

$$y_{ijkl} = \mu + kp_i + gd_j + nhom_k + bo(nhom)_{kl} + e_{ijkl}$$

Trong đó: y_{ijk} : quan sát ở khẩu phần thứ i, giai đoạn j, nhóm bò thứ k và của bò l trong nhóm k,

μ : trung bình chung,

kp_i : ảnh hưởng của khẩu phần i,

gd_j : ảnh hưởng của giai đoạn j,

$nhom_k$: ảnh hưởng của nhóm k

$bo(nhom)_{kl}$: ảnh hưởng ngẫu nhiên của bò l trong nhóm k,

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Trong mô hình này yếu tố thí nghiệm được chia thành 2 nhóm: (1) yếu tố cố định (fixed factor) bao gồm *khẩu phần* (KP), *giai đoạn* (GD) và *nhóm bò* (NHOM) và (2) yếu tố ngẫu nhiên (random factor) *bò* (BO). Vì vậy thủ tục GLM được sử dụng để phân tích ảnh hưởng của các yếu tố thí nghiệm. Thủ tục GLM đã được đề cập ở *phần 2.2.3 mục a*.

SAS CODE của VIDU6b:

```
data WORK.VIDU6 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU6.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat GD best32. ;
  informat KP best32. ;
  informat NHOM best32. ;
  informat BO best32. ;
  informat SLS best32. ;
  format GD best12. ;
  format KP best12. ;
  format NHOM best12. ;
  format BO best12. ;
  format SLS best12. ;
input
  GD
  KP
  NHOM
  BO
  SLS
;
PROC GLM; ❶
CLASS KP GD NHOM BO; ❷
```

```
MODEL SLS = KP GD NHOM BO(NHOM)/SS1; ❸
RANDOM BO(NHOM); ❹
RUN;
```

Trong đó: ❶ Thủ tục GLM được sử dụng để phân tích phương sai khi trong mô hình có cả yếu tố cố định và ngẫu nhiên; ❷ Yếu tố thí nghiệm gồm khẩu phần (KP), giai đoạn (GD), nhóm bò (NHOM) và bò (BO); ❸ Mô hình (MODEL) phân tích ảnh hưởng của các yếu tố thí nghiệm (KP GD NHOM và bò trong từng nhóm BO(NHOM)) đến sản lượng sữa bò (SLS); ❹ Ảnh của bò (BO) trong từng nhóm là ảnh hưởng ngẫu nhiên (RANDOM). ❸/SS1 là phần lựa chọn (option); với lựa chọn này SAS chỉ cho kết quả phân tích phương sai chỉ ra ảnh hưởng của từng yếu tố thí nghiệm; nếu không có lựa chọn /SS1 SAS có thể ước tính thêm các phương sai thành phần (bạn đọc có thể thêm bớt lựa chọn này).

Kết quả từ SAS:

The GLM Procedure ❶

Class Level Information		
Class	Levels	Values
KP	2	1 2
GD	2	1 2
NHOM	2	1 2
BO	10	1 2 3 4 5 6 7 8 9 10

Number of observations	20
------------------------	----

The GLM Procedure

Dependent Variable: SLS

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	1337.600000	121.600000	79.74	<.0001
Error	8	12.200000	1.525000		
Corrected Total	19	1349.800000			

R-Square	Coeff Var	Root MSE	SLS Mean
0.990962	4.130130	1.234909	29.90000

Source	DF	Type I SS	Mean Square	F Value	Pr > F ❷
KP	1	57.800000	57.800000	37.90	0.0003
GD	1	45.000000	45.000000	29.51	0.0006
NHOM	1	16.200000	16.200000	10.62	0.0115
BO(NHOM)	8	1218.600000	152.325000	99.89	<.0001

Trong đó: ❶ Kết quả phân tích phương sai với thủ tục GLM; việc đọc kết quả tương tự như thủ tục ANOVA; ❷ Xác suất của từng yếu tố thí nghiệm. Trong thí nghiệm này, khẩu phần (KP) là yếu tố quan tâm chính; nghĩa là sản lượng sữa bò có sự khác biệt giữa 2 công thức hay không? Xác suất của phép phân tích đối với công thức thức ăn $P = 0,0003 < 0,05$. Như vậy công thức thức ăn có ảnh hưởng đến sản lượng sữa của bò.

Khi sử dụng thủ tục GLM, giá trị trung bình bình phương LSM (Least Square Mean) và sai số chuẩn SE (Standard Error) thường được sử dụng để trình bày kết quả. Ước tính LSM, SE và so sánh cặp có thể được thực hiện bằng câu lệnh LSMEANS trong thủ tục GLM với câu lệnh như sau.

```
PROC GLM;
CLASS KP GD NHOM BO;
MODEL SLS = KP GD NHOM BO(NHOM)/SS1;
RANDOM BO(NHOM);
LSMEANS KP /STDERR PDIFF; ❶
RUN;
```

Trong đó: ❶ LSMEANS để ước tính LSM của từng khẩu phần (KP), /STDERR PDIFF các lựa chọn tương ứng để tính SE và so sánh cặp.

Kết quả của câu lệnh này từ SAS như sau:

Least Squares Means				
KP	❶ SLS LSMEAN	❷ Standard Error	H0:LSMEAN=0	❸ H0:LSMean1=LSMean2
			Pr > t	❹ Pr > t
1	31.6000000	0.3905125	<.0001	0.0003
2	28.2000000	0.3905125	<.0001	

Trong đó: ❶ Giá trị LSM của sản lượng sữa (SLS) đối với khẩu phần 1 và 2; ❷ Giá trị SE; ❸ Giả thiết H_0 hai LSM bằng nhau; ❹ Xác suất đối với giả thiết H_0 . Vì $P = 0,0003 < 0,05$ nên bác bỏ H_0 ; tức là hai giá trị LSM không bằng nhau.

Đối với 2 mô hình phân tích nêu trên, đều có chung kết luận là yếu tố *khẩu phần* có ảnh hưởng đến sản lượng sữa, tuy nhiên xác suất (P) ở 2 mô hình đối với yếu tố *khẩu phần* khác nhau (ở mô hình thứ nhất và thứ hai giá trị P lần lượt là 0,0146 và 0,0003). Như vậy việc sử dụng mô hình phân tích phù hợp sẽ cho kết quả có độ tin cậy cao hơn.

d. Thí nghiệm kiểu ô vuông La tinh

Đối với mô hình thí nghiệm ô vuông La tinh, ngoài yếu tố thí nghiệm chính ta còn 2 yếu tố khác: 1) yếu tố hàng và 2) yếu tố cột.

Ví dụ 2.7a: Một thí nghiệm được tiến hành nhằm xác định ảnh hưởng của các loại thức ăn bổ sung khác nhau (A, B, C và D) đến lượng cỏ khô mà bê nuôi vỗ béo thu nhận được (kg/ngày). Thí nghiệm được thiết kế theo mô hình ô vuông la tinh với 4 động vật trong 4 giai đoạn, mỗi giai đoạn 20 ngày. Trong mỗi giai đoạn 10 ngày đầu được coi là giai đoạn thích nghi, 10 ngày tiếp theo là giai đoạn thí nghiệm để thu thập số liệu. Số liệu thu được là khối lượng cỏ khô trung bình bê thu nhận được ở 10 ngày thí nghiệm. Hãy rút ra kết luận từ thí nghiệm nêu trên.

Giai đoạn	Bê			
	1	2	3	4
1	10,0 (B)	9,0 (D)	11,1 (C)	10,8 (A)
2	10,2 (C)	11,3 (A)	9,5 (D)	11,4 (B)
3	8,5 (D)	11,2 (B)	12,8 (A)	11 (C)
4	11,1 (A)	11,4 (C)	11,7 (B)	9,9 (D)

Mô hình phân tích:

$$y_{ijk} = \mu + h_i + c_j + a_k + e_{ijk}$$

Trong đó: y_{ijk} : quan sát ở hàng thứ i , cột thứ j và ở nghiệm thức k

μ : trung bình chung,

h_i : ảnh hưởng của hàng (giai đoạn) i ,

c_j : ảnh hưởng của cột (bê) j ,

a_k : ảnh hưởng của thức ăn bổ sung k ,

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Số liệu có thể nhập vào file VIDU7a.XLS định dạng excel sau đó chuyển file VIDU7a.TXT định dạng txt với 4 cột lần lượt là giai đoạn (GD), bê (BE), thức ăn bổ sung (TA) và khối lượng cỏ khô thu nhận (KLCO).

SAS CODE của VIDU7a.SAS:

```
data WORK.VIDU7A
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU7a.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat GD best32. ;
informat BE best32. ;
informat TA $1. ;
informat KLCO best32. ;
format GD best12. ;
format BE best12. ;
format TA $1. ;
format KLCO best12. ;
input
GD
BE
TA $
KLCO
;
ODS HTML FILE = "D:\SAS2014\KQVIDU7a.HTML";
PROC ANOVA; ❶
CLASS GD BE TA; ❷
MODEL KLCO = GD BE TA; ❸
MEANS TA / TUKEY; ❹
RUN;
ODS HTML CLOSE;
```

Trong đó: ❶ Thủ tục phân tích phương sai được sử dụng để phân tích dữ liệu; ❷ Yếu tố thí nghiệm (GD BE TA); ❸ Mô hình ảnh hưởng của yếu tố thí nghiệm (GD BE TA) đến lượng cỏ khô thu nhận (KLCO); ❹ So sánh giá trị trung bình theo cặp giữa các nghiệm thức của yếu tố thức ăn (TA) bằng phương pháp TUKEY.

Kết quả từ SAS:

The ANOVA Procedure ❶

Class Level Information		
Class	Levels	Values
GD	4	1 2 3 4

Class Level Information		
Class	Levels	Values
BE	4	1 2 3 4
TA	4	A B C D

Number of observations	16
------------------------	----

The ANOVA Procedure

Dependent Variable: KLCO ②

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	17.09562500	1.89951389	13.12	0.0027
Error	6	0.86875000	0.14479167		
Corrected Total	15	17.96437500			

R-Square	Coeff Var	Root MSE	KLCO Mean
0.951640	3.562458	0.380515	10.68125

Source	DF	Anova SS	Mean Square	F Value	Pr > F ③
GD	3	1.48187500	0.49395833	3.41	0.0938
BE	3	3.59187500	1.19729167	8.27	0.0149
TA	3	12.02187500	4.00729167	27.68	0.0007

Tukey's Studentized Range (HSD) Test for KLCO ④

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	6
Error Mean Square	0.144792
Critical Value of Studentized Range	4.89559
Minimum Significant Difference	0.9314

Means with the same letter
are not significantly different.

Tukey Grouping ❸	Mean	N	TA
A	11.5000	4	A
A	11.0750	4	B
A	10.9250	4	C
B	9.2250	4	D

Trong đó: ❶ Kết quả phân tích phương sai; ❷ Bảng phân tích phương sai ảnh hưởng của các yếu tố thí nghiệm đến khối lượng cỏ khô thu nhận (*KLCO*); ❸ Giá trị xác suất đối với từng yếu tố thí nghiệm tương ứng. Kết quả phân tích cho thấy xác suất của phép thử đối với yếu tố thí nghiệm (TA) $P = 0,0007$, vì vậy giả thiết H_0 bị bác bỏ. Kết luận: *Có ảnh hưởng của thức ăn bổ sung đến lượng cỏ khô thu nhận của bê* ($P < 0,05$); ❹ Kết quả so sánh cặp theo phương pháp Tukey; ❺ Sự sai khác giữa các nghiệm thức được tạo ra.

Lưu ý: Một trong những hạn chế của mô hình ô vuông Latinh là số bậc tự do của sai số ngẫu nhiên (xem bảng phân tích phương sai ở ví dụ 2.7a) rất bé nên các kiểm định F trong phân tích phương sai có thể mắc sai lầm trong kết luận. Việc thiết kế 2 hoặc nhiều ô vuông Latinh cho một thí nghiệm có thể làm tăng độ chính xác của kết luận. Nếu thí nghiệm được tiến hành trên nhiều ô vuông Latinh khác nhau việc phân tích số liệu sẽ bao gồm ảnh hưởng của 3 yếu tố trong một ô vuông (hàng, cột, yếu tố thí nghiệm) và ảnh hưởng của các ô.

Ví dụ 2.7b: Giả sử, một thí nghiệm được thiết kế tương tự như ở ví dụ trên, nhưng có 2 ô vuông Latinh được thiết kế đồng thời và mỗi ô đều có 4 động vật thí nghiệm và 4 công thức thí nghiệm khác nhau. Số liệu ở ô vuông Latinh thứ nhất như trong ví dụ 7A, ô vuông Latinh thứ 2 như trong bảng dưới đây. Hãy tiến hành phân tích để đưa ra kết luận và đưa ra nhận xét về mô hình thiết kế trong 2 ví dụ vừa rồi.

Ô vuông 1					Ô vuông 2				
Giai đoạn	Bê				Giai đoạn	Bê			
	1	2	3	4		1	2	3	4
1	10,0 (B)	9,0 (D)	11,1 (C)	10,8 (A)	1	10,9 (C)	11,2 (A)	9,4 (D)	11,2 (B)
2	10,2 (C)	11,3 (A)	9,5 (D)	11,4 (B)	2	10,5 (B)	9,6 (D)	11,4 (C)	10,9 (A)
3	8,5 (D)	11,2 (B)	12,8 (A)	11 (C)	3	11,1 (A)	11,4 (C)	11,7 (B)	9,8 (D)
4	11,1 (A)	11,4 (C)	11,7 (B)	9,9 (D)	4	8,8 (D)	12,9 (B)	11,4 (A)	11,2 (C)

Mô hình phân tích:

$$y_{ijkm} = \mu + o_m + h(o)_{im} + c(o)_{jm} + a_k + e_{ijkm}$$

Trong đó: y_{ijkm} : quan sát ở hàng thứ i, cột thứ j và ở nghiệm thức k

μ : trung bình chung,

α_m : ảnh hưởng của ô vuông m

$h(\alpha)_{im}$: ảnh hưởng của hàng i trong ô vuông m,

$c(\alpha)_{jm}$: ảnh hưởng của cột j trong ô vuông m,

α_k : ảnh hưởng của mức k của nhân tố,

e_{ijkm} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Ngoài 4 cột lần lượt là giai đoạn (GD), bê (BE), thức ăn bổ sung (TA) và khối lượng cỏ khô thu nhận (KLCO) như ở đối với ví dụ 2.7a, cần bổ sung thêm cột ô vuông (OV) để phân biệt số liệu của 2 ô vuông. Tham khảo cấu trúc số liệu file VIDU7b.XLS định dạng excel và file VIDU7b.TXT.

SAS CODE của VIDU7b.SAS:

```
data WORK.VIDU7b          ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU7B.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
    informat OV best32. ;
    informat GD best32. ;
    informat BE best32. ;
    informat TA $1. ;
    informat KLCO best32. ;
    format OV best12. ;
    format GD best12. ;
    format BE best12. ;
    format TA $1. ;
    format KLCO best12. ;
input
    OV
    GD
    BE
    TA $
    KLCO
;
PROC ANOVA;
CLASS OV GD BE TA;
MODEL KLCO = OV GD(OV) BE(OV) TA;
```

```
MEANS TA /TUKEY;
RUN;
```

Kết quả từ SAS:

The ANOVA Procedure **①**

Class Level Information		
Class	Levels	Values
OV	2	1 2
GD	4	1 2 3 4
BE	4	1 2 3 4
TA	4	A B C D

Number of observations	32 ②
------------------------	-------------

The ANOVA Procedure

Dependent Variable: KLCO

Source	DF ③	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	30.50250000	1.90640625	9.49	<.0001
Error	15	3.01468750	0.20097917		
Corrected Total	31	33.51718750			

R-Square	Coeff Var	Root MSE	KLCO Mean
0.910055	4.166664	0.448307	10.75938

Source	DF ④	Anova SS	Mean Square	F Value	Pr > F ⑤
OV	1	0.19531250	0.19531250	0.97	0.3399
GD(OV)	6	2.14437500	0.35739583	1.78	0.1712
BE(OV)	6	5.49937500	0.91656250	4.56	0.0079
TA	3	22.66343750	7.55447917	37.59	<.0001

Duncan's Multiple Range Test for KLCO **⑥**

Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	0.200979

Number of Means	2	3	4
Critical Range	.4778	.5008	.5152

Means with the same letter are not significantly different.			
Duncan Grouping	Mean	N	TA
A	11.3250	8	A
A	11.3250	8	B
A	11.0750	8	C
B	9.3125	8	D

Trong đó: ❶ Kết quả phân tích phương sai; ❷ Tổng số quan sát trong thí nghiệm $n = 32$; ❸ Bậc tự do (df) tương ứng với từng nguồn biến động: Tổng số (Corrected Total) = $31 = 32 - 1$, Mô hình (Model) = $16 = 1 + 6 + 6 + 3$ (xem ❹), Sai số ngẫu nhiên (Error) = $31 - 16 = 15$; ❺ Xác suất đối với từng yếu tố thí nghiệm; ❻ So sánh cặp bằng phép thử Tukey.

Giải thích: So sánh với ví dụ 2.7a, bậc tự do sai số ngẫu nhiên trong ví dụ này đã được cải thiện đáng kể ($df = 15$). Kết quả phân tích cho thấy xác suất của phép thử đối với yếu tố thí nghiệm (TA) $P < 0,0001$ (xem ❺), vì vậy giả thiết H_0 bị bác bỏ. Kết luận: Có ảnh hưởng của thức bổ sung đến lượng cỏ khô thu nhận của bê.

e. Thí nghiệm 2 yếu tố chéo nhau

Với mô hình thí nghiệm 2 yếu tố chéo nhau, ngoài nghiên cứu tác động của từng yếu tố thí nghiệm chính ta còn nghiên cứu mối tương tác giữa 2 yếu tố này.

Ví dụ 2.8: Một thí nghiệm yếu tố 2×2 được tiến hành để xác định ảnh hưởng của việc bổ sung 2 loại vitamin (A và B) vào thức ăn đến tăng khối lượng (kg/ngày) của lợn. Hai mức đối với vitamin A (0 và 4mg) và 2 mức đối với vitamin B (0 và 5mg) được sử dụng trong thí nghiệm này. Tổng số 20 lợn thí nghiệm được phân về 4 tổ hợp (công thức thí nghiệm) một cách ngẫu nhiên. Số liệu thu được khi kết thúc thí nghiệm được trình bày như sau:

A	0mg		mg	
B	0mg	5mg	0mg	5mg
	0,585	0,567	0,473	0,684
	0,536	0,545	0,450	0,702
	0,458	0,589	0,869	0,900
	0,486	0,536	0,473	0,698
	0,536	0,549	0,464	0,693

Mô hình phân tích:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}$$

Trong đó: y_{ijk} : quan sát thứ k ở mức i của yếu tố A và mức j của yếu tố B

μ : là trung bình chung

α_i : ảnh hưởng mức i của yếu tố A,

β_j : ảnh hưởng mức j của yếu tố B,

$(\alpha\beta)_{ij}$: tương tác giữa mức i của yếu tố A và mức j của yếu tố B

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Số liệu có thể nhập vào file VIDU8.XLS định dạng Excel sau đó chuyển file VIDU8.TXT định dạng txt với 3 cột lần lượt là vitamin A (VITA), vitamin B (VITB) và tăng khối lượng trung bình (TT).

SAS CODE của VIDU8.SAS:

```
data WORK.VIDU8 ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU8.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat VITA best32. ;
informat VITB best32. ;
informat TT best32. ;
format VITA best12. ;
format VITB best12. ;
format TT best12. ;
input
    VITA
    VITB
    TT
;
PROC ANOVA;
CLASS VITA VITB;
MODEL TT = VITA VITB VITA*VITB;
RUN;
```

Kết quả từ SAS:

The ANOVA Procedure ❶

Dependent Variable: TT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.14521095	0.04840365	4.39	0.0196

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Error	16	0.17648360	0.01103023		
Corrected Total	19	0.32169455			

R-Square	Coeff Var	Root MSE	TT Mean
0.451394	17.81139	0.105025	0.589650

Source	DF	Anova SS	Mean Square	F Value	Pr > F ^②
VITA	1	0.05191805	0.05191805	4.71	0.0454
VITB	1	0.06418445	0.06418445	5.82	0.0282
VITA*VITB	1	0.02910845	0.02910845	2.64	0.1238

Trong đó: ① Kết quả phân tích phương sai; ② Xác suất đối với từng yếu tố phân tích. Các giá trị xác suất lần lượt đối với vitamin A (VITA), vitamin B (VITB) và tương tác giữa hai yếu tố (VITA*VITB) là 0,0454; 0,0282 và 0,1238. Với xác suất $P = 0,045$ và $0,028 < 0,05$. Kết luận: Vitamin A và B có ảnh hưởng đến tăng khối lượng của lợn, nhưng không có sự tương tác giữa hai loại vitamin ($P = 0,1238 > 0,05$).

g. Thí nghiệm hai nhân tố phân cấp (nested design)

Với mô hình hai nhân tố phân cấp, nhân tố cấp trên (A) là cố định và cấp dưới (B) là ngẫu nhiên. Hoặc cả hai yếu tố thí nghiệm là ngẫu nhiên. Như vậy B sẽ làm ổ (nested) trong A.

Ví dụ 2.9: Thí nghiệm nhằm xác định ảnh hưởng của lợn đực giống và lợn nái đến khối lượng sơ sinh của thể hệ con. Mô hình phân cấp 2 yếu tố được sử dụng. Bốn lợn đực giống được chọn ngẫu nhiên ($a = 4$), mỗi đực phối với 3 lợn nái ($b = 3$) và mỗi nái sinh được 2 lợn con ($r = 2$). Khối lượng (kg) sơ sinh của từng lợn con thu được như sau:

Đực	<u>1</u>			<u>2</u>			<u>3</u>			<u>4</u>		
Nái	1	2	3	4	5	6	7	8	9	10	11	12
Con	1,2	1,2	1,1	1,2	1,1	1,2	1,2	1,3	1,2	1,3	1,4	1,3
	1,2	1,3	1,2	1,2	1,2	1,1	1,2	1,3	1,2	1,3	1,4	1,3

Trong mô hình phân thí nghiệm này, mỗi lợn đực luôn được ghép đôi giao phối với 3 nái cho trước. Khái niệm này trong thuật ngữ thống kê gọi là mô hình làm tổ (nested).

Mô hình phân tích:

$$y_{ijk} = \mu + a_i + b_{j(i)} + e_{ijk}$$

Trong đó: y_{ijk} : quan sát thứ k ở mức j của yếu tố B trong mức i của yếu tố A;

μ : trung bình chung;

a_i : ảnh hưởng mức thứ i của nhân tố A;

$b_{j(i)}$: ảnh hưởng mức j của yếu tố B trong mức i của yếu tố A;

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Số liệu có thể nhập vào file VIDU9.XLS định dạng excel sau đó chuyển file VIDU9.TXT định dạng txt với 3 cột lần lượt là số hiệu của bố (BO), số hiệu của mẹ (ME) và khối lượng của lợn con (KL).

SAS CODE của VIDU9.SAS:

```
data WORK.VIDU9
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU9.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat BO best32. ;
informat ME best32. ;
informat KL best32. ;
format BO best12. ;
format ME best12. ;
format KL best12. ;
input
BO
ME
KL
;
PROC SORT;❶
BY BO ME;
RUN;

PROC NESTED;❷
CLASS BO ME;
VAR KL;
RUN;

PROC MIXED;❸
CLASS BO ME;
MODEL KL = BO;
RANDOM ME(BO);
RUN;
```

Trong đó: ❶ Câu lệnh PROC SORT được sử dụng trước; ❷ câu lệnh PROC NESTED nếu cả 2 yếu tố đều là ngẫu nhiên và yếu tố đứng sau trong câu lệnh CLASS được mặc định là nested (làm tổ) trong yếu tố đứng trước. Nếu trong mô hình có cả yếu tố ngẫu nhiên (ME) và cố định (BO); ❸ câu lệnh PROC MIXED là một giải pháp phù hợp.

Kết quả từ SAS: 2 output (Nested Procedure và Mixed Procedure).

The NESTED Procedure ❶								
Coefficients of ❷ Expected Mean Squares								
Source	BO	ME	Error					
BO	6	2	1					
ME	0	2	1					
Error	0	0	1					

Nested Random Effects Analysis of Variance for Variable KL								
Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Square	Variance Component ❸	Percent of Total
Total	23	0.153333				0.006667	0.007685	100.0000
BO	3	0.093333	6.22	0.0174	ME	0.031111	0.004352	56.6265
ME	8	0.040000	3.00	0.0424	Error	0.005000	0.001667	21.6867
Error	12	0.020000				0.001667	0.001667	21.6867

KL Mean	1.23333333
Standard Error of KL Mean	0.03600411

Trong đó: ❶ Kết quả phân tích theo câu lệnh PROC NESTED. Kết quả từ câu lệnh này được sử dụng để tính phương sai thành phần nhằm tính được hệ số di truyền (tham khảo tài liệu về Di truyền số lượng); ❷ Hệ số để ước tính các trung bình bình phương - phương sai thành phần E(MS), (xem chi tiết trang 86 Giáo trình Thiết kế thí nghiệm của Nguyễn Đình Hiền và Đỗ Đức Lực, 2007); ❸ Các phương sai thành phần được tính dựa vào các hệ số đã nêu. Ví dụ:

$$E(MS_{Error}) = 0,001667 = 1 * MS_{Error} \Rightarrow MS_{Error} = 0,001667/1 = 0,001667;$$

$$E(MS_{ME}) = 0,005000 = 2 * MS_{ME} + 1 * MS_{Error} = 2 * MS_{ME} + 1 * 0,001667$$

$$\Rightarrow MS_{ME} = (0,005000 - 0,001667)/2 = 0,001667;$$

$$\begin{aligned} E(MS_{BO}) &= 0,031111 = 6 * MS_{BO} + 2 * MS_{ME} + 1 * MS_{Error} \\ &= 6 * MS_{BO} + 2 * 0,001667 + 1 * 0,001667 \end{aligned}$$

$$\Rightarrow MS_{BO} = (0,031111 - 2*0,001667 - 1*0,001667)/6 = 0,004352$$

The Mixed Procedure **1**

Model Information	
Data Set	WORK.VIDU9
Dependent Variable	KL
Covariance Structure	Variance Components
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	Levels	Values
BO	4	1 2 3 4
ME	12	1 2 3 4 5 6 7 8 9 10 11 12

Dimensions	
Covariance Parameters	2
Columns in X	5
Columns in Z	12
Subjects	1
Max Obs Per Subject	24
Observations Used	24
Observations Not Used	0
Total Observations	24

Iteration History			
Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-52.25828060	
1	1	-55.22511559	0.00000000

Convergence criteria met.

Covariance Parameter Estimates	
Cov Parm	Estimate
ME(BO)	0.001667
Residual	0.001667

Fit Statistics	
-2 Res Log Likelihood	-55.2
AIC (smaller is better)	-51.2
AICC (smaller is better)	-50.5
BIC (smaller is better)	-50.3

Type 3 Tests of Fixed Effects ²				
Effect	Num DF	Den DF	F Value	Pr > F
BO	3	8	6.22	0.0174

Trong đó: ❶ Kết quả phân tích theo câu lệnh PROC MIXED. Trong mô hình này bao gồm yếu tố cố định và yếu tố ngẫu nhiên. Nếu bạn đọc quan tâm đến ảnh hưởng của yếu tố cố định trong mô hình mixed, câu lệnh PROC MIXED sẽ cho ❷ kết quả phân tích với xác suất đối với yếu tố cố định $P = 0,017$. Nếu tất cả các yếu tố thí nghiệm đều là ngẫu nhiên trong mô hình phân tích, có thể sử dụng PROC GLM để phân tích dữ liệu.

h. Thí nghiệm 2 nhân tố chia ô

Ví dụ 2.10: Một thí nghiệm được tiến hành để nghiên cứu ảnh hưởng của bãi chăn thả A (1, 2, 3 và 4) và lượng khoáng bổ sung B (1 và 2) đến năng suất sữa. Có tất cả 24 bò tham gia thí nghiệm. Thí nghiệm được thiết kế theo mô hình hai nhân tố kiểu chia ô với yếu tố A được bố trí trên ô lớn và yếu tố B trên ô nhỏ trên 3 khối. Năng suất sữa trung bình được ghi lại như sau (kg/ngày):

Khối 1			
A ₄	A ₁	A ₂	A ₃
B ₂	B ₂	B ₁	B ₂
30	27	26	26
B ₁	B ₁	B ₂	B ₁

Khối 2			
A ₂	A ₁	A ₄	A ₃
B ₁	B ₂	B ₁	B ₁
32	30	34	33
B ₂	B ₁	B ₂	B ₂

Khối 3			
A ₁	A ₂	A ₄	A ₃
B ₂	B ₁	B ₂	B ₁
34	30	36	33
B ₁	B ₂	B ₁	B ₂

29	25	28	24
----	----	----	----

37	31	37	
----	----	----	--

--	--	--	--

Mô hình phân tích:

$$y_{ijl} = \mu + a_i + k_l + (ak)_{il} + b_j + (ab)_{ij} + e_{ijl}$$

Trong đó: μ : trung bình chung

a_i : ảnh hưởng của mức i của nhân tố A (trên ô lớn);

b_j : ảnh hưởng của mức j của nhân tố B (trên ô nhỏ);

k_l : ảnh hưởng của khối l;

$(ak)_{il}$: tương tác giữa nhân tố A và khối được dùng làm sai số ô lớn se^2_L ;

$(ab)_{ij}$: tương tác của hai nhân tố A và B;

e_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$.

Lưu ý: Trong mô hình này khối coi như nhân tố ngẫu nhiên, không tương tác với B, hai nhân tố A và B coi như nhân tố cố định.

Số liệu có thể nhập vào file VIDU10.XLS định dạng excel sau đó chuyển file VIDU10.TXT định dạng txt với 4 cột lần lượt là khối (KHOI), bãi chăn thả (A), lượng khoáng bổ sung (B) và sản lượng sữa (SLS).

SAS CODE của VIDU10.SAS:

```
data WORK.VIDU10
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU10.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat KHOI best32. ;
informat A best32. ;
informat B best32. ;
informat SLS best32. ;
format KHOI best12. ;
format A best12. ;
format B best12. ;
format SLS best12. ;
input
    KHOI
    A
    B
    SLS
;
;
;
```

```

PROC GLM; ❶

CLASS KHOI A B; ❷

MODEL SLS = KHOI A KHOI*A B A*B /SS1; ❸

RANDOM KHOI KHOI*A; ❹

LSMEANS A B / STDERR PDIF; ❺

RUN;

```

Trong đó: ❶ Thủ tục PROC GLM sử dụng để phân tích số liệu với ❷ các yếu tố thí nghiệm (KHOI A B) và ❸ Mô hình phân tích ảnh hưởng của yếu tố thí nghiệm đến sản lượng sữa trong đó ❹ KHOI KHOI*A là yếu tố ngẫu nhiên; ❺ Ước tính giá trị LSM và SE đối với các nghiệm thức của yếu tố A và B, so sánh các LSM theo cặp bằng lệnh PDIF.

Kết quả từ SAS:

The GLM Procedure❶

Class Level Information❷		
Class	Levels	Values
KHOI	3	1 2 3
A	4	1 2 3 4
B	2	1 2

Number of observations	24
------------------------	----

The GLM Procedure

Dependent Variable: SLS❸

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	323.8333333	21.5888889	9.60	0.0015
Error	8	18.0000000	2.2500000		
Corrected Total	23	341.8333333			

R-Square	Coeff Var	Root MSE	SLS Mean
0.947343	4.825737	1.500000	31.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F❹
KHOI	2	212.5833333	106.2916667	47.24	<.0001
A	3	71.1666667	23.7222222	10.54	0.0037

Source	DF	Type I SS	Mean Square	F Value	Pr > F ^④
KHOI*A	6	26.0833333	4.3472222	1.93	0.1909
B	1	8.1666667	8.1666667	3.63	0.0932
A*B	3	5.8333333	1.9444444	0.86	0.4981

The GLM Procedure ^⑤

Least Squares Means

A ^⑥	SLS LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	29.6666667	0.6123724	<.0001	1
2	30.6666667	0.6123724	<.0001	2
3	30.0000000	0.6123724	<.0001	3
4	34.0000000	0.6123724	<.0001	4

Least Squares Means for effect A ^⑦ Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: SLS				
i/j	1	2	3	4
1		0.2815	0.7103	0.0010
2	0.2815		0.4635	0.0049
3	0.7103	0.4635		0.0017
4	0.0010	0.0049	0.0017	

B ^⑧	SLS LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2
			Pr > t	Pr > t ^⑨
1	30.5000000	0.4330127	<.0001	0.0932
2	31.6666667	0.4330127	<.0001	

Trong đó: ^① Kết quả phân tích bằng PROC GLM; ^② Thông tin về yếu tố thí nghiệm (KHOI, A và B); ^③ Kết quả phân tích phương sai đối với biến sản lượng sữa (SLS); ^④ Xác suất đối với các yếu tố trong mô hình phân tích. Trong kết quả phân tích ta chỉ quan tâm đến xác suất P đối với yếu tố A, B và tương tác A*B. Các giá trị này lần lượt là 0,0377; 0,0932 và 0,4981. Với các giá trị này ta có thể kết luận năng suất sữa có sự khác nhau giữa các bãi chăn thả ($P < 0,05$). Tuy nhiên, việc bổ sung các khoáng chất không làm ảnh hưởng đến năng suất sữa và cũng không có ảnh hưởng tương tác giữa bãi chăn thả và việc bổ sung khoáng ($P > 0,05$); ^⑤ Kết quả ước tính giá trị LSM và SE

đối với ❹ Yếu tố A và ❸ Yếu tố B và so sánh cặp giữa các LSM của ❹ Yếu tố A và ❸ Yếu tố B.

Ví dụ 2.10B: Xem xét ví dụ trên, giả sử rằng thí nghiệm được thực hiện không có khối và chỉ có yếu tố A và B được thiết kế trên 12 ô lớn. Năng suất sữa trung bình như sau (kg/ngày):

1	2	3	4	5	6	7	8	9	10	11	12
A ₄	A ₁	A ₂	A ₃	A ₂	A ₁	A ₄	A ₃	A ₁	A ₂	A ₄	A ₃
B ₂ 30	B ₂ 27	B ₁ 26	B ₂ 26	B ₁ 32	B ₂ 30	B ₁ 34	B ₁ 33	B ₂ 34	B ₁ 30	B ₂ 36	B ₁ 33
B ₁ 29	B ₁ 25	B ₂ 28	B ₁ 24	B ₂ 37	B ₁ 31	B ₂ 37	B ₂ 32	B ₁ 31	B ₂ 31	B ₁ 38	B ₂ 32

Mô hình phân tích:

$$y_{ijkl} = \mu + a_i + o(a)_{ik} + b_j + (ab)_{ij} + e_{ijkl}$$

Trong đó: μ : trung bình chung

a_i : ảnh hưởng của mức i của nhân tố A (trên ô lớn);

b_j : ảnh hưởng của mức j của nhân tố B (trên ô nhỏ);

$o(a)_{ik}$: sai số của ô lớn;

$(ab)_{ij}$: tương tác của hai nhân tố A và B;

e_{ijkl} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$.

Số liệu có thể nhập vào file VIDU10B.XLS định dạng excel sau đó chuyển file VIDU10B.TXT định dạng txt với 4 cột lần lượt là ô (O), bãi chăn thả (A), lượng khoáng bổ sung (B) và sản lượng sữa (SLS).

SAS CODE của VIDU10B.SAS:

```
data WORK.VIDU10B
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU10B.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat O best32. ;
informat A best32. ;
informat B best32. ;
informat SLS best32. ;
format O best12. ;
```

```

format A best12. ;
format B best12. ;
format SLS best12. ;

input
    O
    A
    B
    SLS
;
PROC GLM;
CLASS O A B;
MODEL SLS = A B A*B O(A) / SS1;
RANDOM O(A);
LSMEANS A B / STDERR PDIF;
RUN;

```

Kết quả từ SAS:

The GLM Procedure ❶

Class Level Information		
Class	Levels	Values
O	12	1 2 3 4 5 6 7 8 9 10 11 12
A	4	1 2 3 4
B	2	1 2

Number of observations	24
------------------------	----

The GLM Procedure

Dependent Variable: SLS ❷

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	323.8333333	21.58888889	9.60	0.0015
Error ❸	8	18.0000000	2.2500000		
Corrected Total	23	341.8333333			

R-Square	Coeff Var	Root MSE	SLS Mean
----------	-----------	----------	----------

R-Square	Coeff Var	Root MSE	SLS Mean
0.947343	4.825737	1.500000	31.08333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	3	71.1666667	23.7222222	10.54	0.0037 ④
B	1	8.1666667	8.1666667	3.63	0.0932 ⑤
A*B	3	5.8333333	1.9444444	0.86	0.4981 ⑥
O(A)	8	238.6666667	29.8333333	13.26	0.0007

Trong đó: **①** Kết quả của thủ tục PROC GLM; **②** Kết quả phân tích phương sai đối với biến sản lượng sữa (*SLS*); **③** Sai số ngẫu nhiên. Trong ví dụ này, yếu tố ô (O) được coi là ngẫu nhiên và làm tổ (nested) trong yếu tố A. Vì vậy sai số của ô lớn (A) chính là O (A). Do đó **④** giá trị F của yếu tố A không phải là $10.54 = MS_A / MS_{Error} = 23,7222222 / 2,2500000$ mà được tính $0,8 = MS_A / MS_{O(A)} = 23,7222222 / 29,8333333 = 0,7952$. Ta sẽ có xác suất đối với yếu tố A là $P = 0,5302$ (bạn đọc có thể tự tính trong Excel với câu lệnh $= 1 - F.DIST(0.7952;3;8;TRUE)$, trong đó 3 và 8 lần lượt là bậc tự do của yếu tố A và của O(A)). Ba giá trị xác suất P quan tâm đến bao gồm 0,5302; **⑤**0,0932 và **⑥**0,4981 tương ứng với yếu tố A, B và tương tác A*B. Với cách thiết kế thí nghiệm theo mô hình thứ 2 này ta đã không tìm thấy ảnh hưởng của bất kỳ một yếu tố nào ($P > 0,05$).

The GLM Procedure

Least Squares Means **①**

A	SLS LSMEAN	Standard Error	Pr > t	LSMEAN Number
1	29.6666667	0.6123724	<.0001	1
2	30.6666667	0.6123724	<.0001	2
3	30.0000000	0.6123724	<.0001	3
4	34.0000000	0.6123724	<.0001	4

Least Squares Means for effect A ② Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: SLS				
i/j	1	2	3	4
1		0.2815	0.7103	0.0010
2	0.2815		0.4635	0.0049

Least Squares Means for effect A ❷ Pr > t for H0: LSMean(i)=LSMean(j) Dependent Variable: SLS				
i/j	1	2	3	4
3	0.7103	0.4635		0.0017
4	0.0010	0.0049	0.0017	

B ❸	SLS LSMEAN	Standard Error	H0:LSMEAN=0	H0:LSMean1=LSMean2
			Pr > t	Pr > t
1	30.5000000	0.4330127	<.0001	0.0932
2	31.6666667	0.4330127	<.0001	

Trong đó: ❶ Giá trị LSM của các yếu tố A và B; ❷ So sánh cặp giữa các nghiệm thức của yếu tố A và ❸ yếu tố B.

h. Thí nghiệm phép đo lặp lại (repeated measures)

Trong các thí nghiệm này, phép đo được lặp lại trên cùng một đơn vị thí nghiệm trong một khoảng thời gian nhất định. Ví dụ, năng suất trong một chu kỳ tiết sữa, sinh trưởng tích lũy của vật nuôi qua các thời điểm khác nhau; giá trị pH, màu sắc của một sản phẩm đo tại các thời điểm khác nhau. Khi tiến hành các phép đo lặp lại như vậy trên cùng một đơn vị thí nghiệm có thể tồn tại mối tương quan giữa các lần đo. Ví dụ, sản lượng sữa bò của một cá thể cao ở tháng thứ 3 thì xu hướng ở tháng thứ 4 cũng sẽ cao mặc dù đó là công thức thí nghiệm nào.

Ví dụ 2.11: Tiến hành đo pH cơ thẩn trên 13 bò vàng, 14 bò LaiSind tại các thời điểm 1, 12, 36, 48 giờ, 6 và 8 ngày (mỗi mẫu được đo lặp lại tại 6 thời điểm khác nhau). Số liệu thu được như sau:

STT*	THOIDIEM	SOTAI	GIONG	pH	STT	THOIDIEM	SOTAI	GIONG	pH
1	01H	1	BV	6.72	26	01H	32	LS	6.71
2	01H	2	BV	6.42	27	01H	33	LS	6.81
3	01H	3	BV	6.51	28	01H	34	LS	6.62
4	01H	4	BV	6.92	29	01H	38	LS	6.81
5	01H	5	BV	6.63	30	01H	39	LS	6.82
6	12H	1	BV	5.39	31	12H	32	LS	5.99
7	12H	2	BV	6.25	32	12H	33	LS	5.98
8	12H	3	BV	5.88	33	12H	34	LS	5.9
9	12H	4	BV	6.92	34	12H	38	LS	5.57
10	12H	5	BV	5.93	35	12H	39	LS	5.54
11	36H	1	BV	5.39	36	36H	32	LS	5.47
12	36H	2	BV	5.47	37	36H	33	LS	5.68
13	36H	3	BV	5.47	38	36H	34	LS	5.46

STT*	THOIDIEM	SOTAI	GIONG	pH	STT	THOIDIEM	SOTAI	GIONG	pH
14	36H	4	BV	5.42	39	36H	38	LS	5.55
15	36H	5	BV	5.43	40	36H	39	LS	5.62
16	48H	1	BV	5.42	41	48H	32	LS	5.45
17	48H	2	BV	5.48	42	48H	33	LS	5.74
18	48H	3	BV	5.48	43	48H	34	LS	5.52
19	48H	4	BV	5.37	44	48H	38	LS	5.56
20	48H	5	BV	5.44	45	48H	39	LS	5.48
21	8D	1	BV	5.43	46	8D	32	LS	5.42
22	8D	2	BV	5.49	47	8D	33	LS	5.77
23	8D	3	BV	5.47	48	8D	34	LS	5.49
24	8D	4	BV	5.4	49	8D	38	LS	5.13
25	8D	5	BV	5.43	50	8D	39	LS	5.55

*: STT = số thứ tự, THOIDIEM = thời điểm sau giết thịt, SOTAI = số tai, GIONG = giống, pH = giá trị pH

Xác định mức độ ảnh hưởng của các yếu tố đến các chỉ tiêu chất lượng thịt theo mô hình thống kê:

$$y_{ijk} = \mu + \alpha_i + \beta_j(\alpha_i) + \tau_k + (\alpha * \tau)_{ik} + \varepsilon_{ijk}$$

Trong đó: y_{ijk} : giá trị quan sát ở thời điểm k đối với động vật thứ j của giống i,

μ : trung bình của chỉ tiêu nghiên cứu,

α_i : ảnh hưởng cố định của giống thứ i,

$\beta_j(\alpha_i)$: ảnh hưởng ngẫu nhiên của động vật thứ j ở giống thứ i,

τ_k : ảnh hưởng của thời điểm sau giết thịt thứ k,

$(\alpha * \tau)_{ik}$: tương tác của giống thứ i với thời điểm sau giết thịt thứ k,

ε_{ijk} : sai số ngẫu nhiên ở thời điểm k đối với động vật thứ j ở giống i.

SAS CODE:

```
data WORK.VIDU11
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU11.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat STT best32. ;
informat THOIDIEM $3. ;
informat SOTAI best32. ;
informat GIONG $2. ;
informat pH best32. ;
format STT best12. ;
format THOIDIEM $3. ;
format SOTAI best12. ;
format GIONG $2. ;
```

```

format pH best12. ;
input
    STT
    THOIDIEM $
    SOTAI
    GIONG $
    pH
;
PROC GLM; ❶
CLASS GIONG SOTAI THOIDIEM; ❷
MODEL pH = GIONG SOTAI(GIONG) THOIDIEM GIONG*THOIDIEM /SS4; ❸
RANDOM SOTAI(GIONG)/TEST; ❹
LSMEANS THOIDIEM/ STDERR PDIFF ADJUST = TUKEY; ❺
RUN;

```

Trong đó: ❶ Thủ tục GLM; ❷ Các yếu tố trong thí nghiệm (GIONG SOTAI THOIDIEM); ❸ Mô hình phân tích ảnh hưởng của các yếu tố đến giá trị pH thịt; ❹ Trong đó số tai bò trong từng giống SOTAI(GIONG) là yếu tố ngẫu nhiên; ❺ Ước tính giá trị LSM, SE và so sánh cặp đôi bằng phép thử Tukey hiệu chỉnh.

Kết quả từ SAS:

The GLM Procedure ❶

Class Level Information ❷		
Class	Levels	Values
GIONG	2	BV LS
SOTAI	10	1 2 3 4 5 32 33 34 38 39
THOIDIEM	5	01H 12H 36H 48H 8D

Number of observations	50
------------------------	----

Dependent Variable: pH ❸

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	17	12.05487200	0.70911012	15.55	<.0001
Error	32	1.45972800	0.04561650		
Corrected Total	49	13.51460000			

R-Square	Coeff Var	Root MSE	pH Mean
0.891989	3.672287	0.213580	5.816000

Source ^④	DF	Type IV SS	Mean Square	F Value	Pr > F
GIONG	1	0.00460800	0.00460800	0.10	0.7527 ^⑤
SOTAI(GIONG)	8	0.51415200	0.06426900	1.41	0.2306
THOIDIEM	4	11.24570000	2.81142500	61.63	<.0001 ^⑥
GIONG*THOIDIEM	4	0.29041200	0.07260300	1.59	0.2005

Source	Type IV Expected Mean Square ^⑦
GIONG	Var(Error) + 5 Var(SOTAI(GIONG)) + Q(GIONG,GIONG*THOIDIEM)
SOTAI(GIONG)	Var(Error) + 5 Var(SOTAI(GIONG))
THOIDIEM	Var(Error) + Q(THOIDIEM,GIONG*THOIDIEM)
GIONG*THOIDIEM	Var(Error) + Q(GIONG*THOIDIEM)

Trong đó: ① Kết quả phân tích từ thủ tục PROC GLM; ② Thông tin về các yếu tố thí nghiệm và các mức tương ứng của từng yếu tố thí nghiệm; ③ Bảng phân tích phương sai của biến phụ thuộc pH; ④ Mức ảnh hưởng của các yếu tố; ⑤ Giá trị pH không có sự sai khác giữa các giống ($P = 0,7527$); ⑥ Giá trị pH thay đổi theo thời gian bảo quản ($P < 0,0001$); ⑦ Bảng ước tính các trung bình bình phương và phương sai thành phần.

Tests of Hypotheses for Mixed Model Analysis of Variance^①

Dependent Variable: pH

	Source	DF	Type IV SS	Mean Square	F Value	Pr > F
*	GIONG	1	0.004608	0.004608	0.07	0.7957
	Error	8	0.514152	0.064269		
Error: MS(SOTAI(GIONG))						
* This test assumes one or more other fixed effects are zero.						

	Source	DF	Type IV SS	Mean Square	F Value	Pr > F
	SOTAI(GIONG)	8	0.514152	0.064269	1.41	0.2306
*	THOIDIEM	4	11.245700	2.811425	61.63	<.0001
	GIONG*THOIDIEM	4	0.290412	0.072603	1.59	0.2005
	Error: MS(Error)	32	1.459728	0.045616		

	Source	DF	Type IV SS	Mean Square	F Value	Pr > F
* This test assumes one or more other fixed effects are zero.						

Trong đó: ❶ Kiểm tra giả thiết đối với phân tích phương sai theo mô hình mixed.

Least Squares Means ❶

Adjustment for Multiple Comparisons: Tukey

THOIDIEM ❷	pH LSMEAN ❸	Standard Error ❹	Pr > t	LSMEAN Number
01H	6.69700000	0.06753999	<.0001	1
12H	5.93500000	0.06753999	<.0001	2
36H	5.49600000	0.06753999	<.0001	3
48H	5.49400000	0.06753999	<.0001	4
8D	5.45800000	0.06753999	<.0001	5

Least Squares Means for effect THOIDIEM Pr > t for H0: LSMean(i)=LSMean(j) ❺ Dependent Variable: pH					
i/j	1	2	3	4	5
1		<.0001	<.0001	<.0001	<.0001
2	<.0001		0.0006	0.0005	0.0002
3	<.0001	0.0006		1.0000	0.9944
4	<.0001	0.0005	1.0000		0.9955
5	<.0001	0.0002	0.9944	0.9955	

Trong đó: ❶ Ước tính giá trị LSM và so sánh cặp theo phương pháp Tukey hiệu chỉnh; ❷ Các thời điểm nghiên cứu; ❸ Giá trị LSM của pH tại các thời điểm tương ứng; ❹ Giá trị SE tương ứng với từng LSM; ❺ Xác suất so sánh cặp theo phương pháp Tukey hiệu chỉnh.

Ngoài thủ tục GLM, thủ tục MIXED cũng có thể được sử dụng để phân tích đối với bài toán phép đo lặp lại.

2.2.4. Phân tích hiệp phương sai (ANCOVA)

Sau đây là một số ví dụ sử dụng SAS để phân tích mô hình thống kê mà biến độc lập có cả biến phân loại và biến liên tục (hiệp biến), tức là phân tích hiệp phương sai (ANCOVA). Trong các nghiên cứu chăn nuôi, thời điểm thu thập dữ liệu khó có thể cố định vào một thời điểm nhất định. Ví dụ: khối lượng của động vật thí nghiệm lúc bắt

đầu thí nghiệm không đồng đều nhau hoặc thiết kế thí nghiệm cân khối lượng của lợn ở thời điểm cai sữa 28 ngày, 60 ngày và kết thúc thí nghiệm lúc 7,5 tháng tuổi. Tuy nhiên, trong thực tế số liệu không đúng thời điểm thiết kế và thường chỉ được thu thập vào các thời điểm xung quanh so với dự kiến. Như vậy khối lượng của động vật thí nghiệm sẽ thấp hơn so với dự kiến nếu tiến hành cân sớm và ngược lại. Để khắc phục những hạn chế nêu trên, khối lượng của động vật thí nghiệm lúc bắt đầu thí nghiệm, thời gian cân thực tế có thể sử dụng như một hiệp biến, tức là một yếu tố gây biến động.

Ví dụ 2.12: Tiến hành nuôi vỗ béo lợn ở 3 công thức thức ăn (A, B và C) trong 90 ngày. Khối lượng (kg) của từng động vật thí nghiệm tại thời điểm bắt đầu (P0) và kết thúc nuôi vỗ béo (P1) và tăng khối lượng (TT, g/ngày) được trình bày ở bảng sau:

KP	A	A	A	A	A	B	B	B	B	B	C	C	C	C	C
P0	35	40	36	35	34	39	34	41	43	39	40	32	33	39	42
P1	122	130	124	123	121	128	120	129	132	127	129	116	117	129	132
TT	967	1000	978	978	967	989	956	978	989	978	989	933	933	1000	1000

Ta có thể dùng thủ tục ANOVA để phân tích phương sai đối với số liệu như mô hình thiết kế thí nghiệm một yếu tố hoàn toàn ngẫu nhiên với biến phụ thuộc là tăng khối lượng và một biến độc lập (yếu tố thí nghiệm) là công thức thức ăn.

Mô hình phân tích

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

Trong đó: y_{ij} : quan sát ở cá thể thứ j ở công thức thức ăn i,

μ : trung bình chung,

α_i : ảnh hưởng của công thức thức ăn i,

e_{ij} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$.

Số liệu có thể nhập vào file VIDU12.XLS định dạng excel sau đó chuyển file VIDU12.TXT định dạng txt với 4 cột lần lượt là khẩu phần (KP), khối lượng lúc bắt đầu thí nghiệm (P0), khối lượng kết thúc thí nghiệm (P1) và tăng khối lượng trung bình/ngày trong thời gian nuôi thí nghiệm (TT).

SAS CODE của VIDU12.SAS:

```
data WORK.VIDU12
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU12.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat KP $1. ;
informat P0 best32. ;
informat P1 best32. ;
informat TT best32. ;
format KP $1. ;
format P0 best12. ;
format P1 best12. ;
```

```

format TT best12. ;
input
    KP $
    P0
    P1
    TT
;
PROC ANOVA;
CLASS KP;
MODEL TT = KP;
RUN;

```

Kết quả từ SAS:

The ANOVA Procedure ^❶

Class Level Information		
Class	Levels	Values
KP	3	A B C

Number of observations	15
------------------------	----

Dependent Variable: TT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	163.333333	81.666667	0.15	0.8586
Error	12	6346.000000	528.833333		
Corrected Total	14	6509.333333			

R-Square	Coeff Var	Root MSE	TT Mean
0.025092	2.356991	22.99638	975.6667

Source	DF	Anova SS	Mean Square	F Value	Pr > F ^❷
KP	2	163.3333333	81.6666667	0.15	0.8586

Trong đó: ^❶ Kết quả phân tích phương sai; ^❷ Xác suất đối với yếu tố thí nghiệm (KP) $P = 0,8586$. Với xác suất này giả thuyết H_0 được chấp nhận, tức là “khẩu phần” không ảnh hưởng đến tăng khối lượng trung bình. Nếu phân tích số liệu theo mô hình nêu trên, ta có thể đã bỏ qua một thông tin quan trọng, đó là những cá thể có khối lượng ban đầu khác nhau có thể cho tăng khối lượng khác nhau. Để khắc phục hạn chế này, khối lượng ban đầu được đưa vào mô hình như hiệp biến và mô hình phân tích hiệp phương sai được sử dụng để phân tích.

Mô hình phân tích:

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + e_{ij}$$

Trong đó: y_{ij} : quan sát ở cá thể thứ j ở công thức thức ăn i,

μ : trung bình chung,

α_i : ảnh hưởng của công thức thức ăn i,

βx_{ij} : ảnh hưởng của khối lượng bắt đầu của cá thể j ở công thức i (hiệp biến),

e_{ij} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$.

Sử dụng lại dữ liệu của ví dụ 2.12 nhưng với câu lệnh SAS như sau:

SAS CODE:

```
PROC GLM; ❶  
CLASS KP; ❷  
MODEL TT = P0 KP; ❸  
RUN;
```

Trong đó: ❶ Sử dụng thực tục GLM để phân tích hiệp phương sai; ❷ yếu tố thí nghiệm là khẩu phần (KP) với ❸ mô hình phân tích hiệp phương sai với hiệp biến là khối bắt đầu (P0) và yếu tố thí nghiệm là khẩu phần (KP).

Kết quả từ SAS:

The GLM Procedure ❶

Dependent Variable: TT ❷

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5409.734432	1803.244811	18.04	0.0001
Error	11	1099.598901	99.963536		
Corrected Total	14	6509.333333			

R-Square	Coeff Var	Root MSE	TT Mean
0.831073 ❸	1.024753	9.998177	975.6667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
P0	1	4543.542961	4543.542961	45.45	<.0001 ❹

Source	DF	Type I SS	Mean Square	F Value	Pr > F
KP	2	866.191472	433.095736	4.33	0.0410 ❸

Trong đó: ❶ Kết quả phân tích hiệp phương sai bằng thủ tục GLM; ❷ Bảng phân tích hiệp phương sai; ❸ Hệ số xác định của mô hình $R^2 = 0,831073$; ❹ Xác suất đối với hiệp phương sai – khối lượng bắt đầu thí nghiệm (P0) $P < 0,0001$, có nghĩa khối lượng lúc bắt đầu thí nghiệm có ảnh hưởng đến tăng khối lượng; ❺ Xác suất đối với yếu tố thí nghiệm - khẩu phần (KP) $P = 0,0410 < 0,05$, có nghĩa khẩu phần có ảnh hưởng đến tăng khối lượng.

Như vậy, áp dụng GLM để phân tích ANCOVA đã cho ta kết quả khác với phân tích ANOVA với một yếu tố hoàn toàn ngẫu nhiên. Trong bảng phân tích phương sai trên ta thấy khối lượng ban đầu ảnh hưởng một cách rõ rệt đến tăng khối lượng của động vật thí nghiệm ($P < 0,001$) và cũng tồn tại ảnh hưởng của các công thức thức ăn đến tăng khối lượng của lợn ($P = 0,0410$); ngược lại ở kết quả phân tích ANOVA cho thấy yếu tố thí nghiệm (khẩu phần) không ảnh hưởng đến tăng khối lượng của vật nuôi ($P = 0,8586$). Bên cạnh đó hệ số xác định của phân tích ANOVA bé hơn nhiều so với ANCOVA và lần lượt là 0,025092 và 0,831073.

2.3. TƯƠNG QUAN VÀ PHÂN TÍCH HỒI QUY

Tương quan được áp dụng đối với các biến định lượng và thể hiện mối liên hệ giữa 2 biến. Phân tích hồi quy được áp dụng khi cả biến phụ thuộc (số liệu thí nghiệm) có dạng biến liên tục và biến độc lập (nhân tố ảnh hưởng) cũng thuộc dạng biến liên tục.

2.3.1. Tương quan

Dựa trên lý thuyết xác suất về hệ số tương quan chúng ta có công thức sau để tính hệ số tương quan mẫu r_{XY} giữa hai biến ngẫu nhiên X và Y:

$$r_{XY} = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2}} = \frac{Cov_{XY}}{S_X S_Y}$$

Ví dụ 2.13: Tiến hành cân khối lượng (P), đo đường kính lớn (D) và đường kính bé (d) của 22 quả trứng gà. Số liệu thu được trình bày ở bảng dưới đây.

P (g)	66,80	60,10	71,20	61,60	61,20	59,00	67,90	59,00	51,50	62,60	64,20
D (mm)	58,37	54,95	60,58	56,73	57,36	53,26	57,07	58,17	52,28	55,62	56,82

d (mm)	45,12	44,35	45,56	44,34	43,57	44,86	46,27	42,82	41,91	44,95	44,79
P (g)	71,20	54,20	54,50	69,10	55,90	66,00	68,00	62,00	56,70	67,00	53,80
D (mm)	61,15	54,24	54,99	60,99	54,41	58,19	59,93	56,80	55,66	58,49	52,44
d (mm)	46,00	42,58	42,32	44,85	42,62	45,69	45,50	44,20	42,41	45,56	43,38

Số liệu có thể nhập vào file VIDU13.XLS định dạng excel sau đó chuyển file VIDU13.TXT định dạng txt với 3 cột lần lượt là khối lượng trứng (KL), đường kính lớn (DL) và đường kính bé (DN).

SAS CODE của VIDU13.SAS sau đây ta có thể tính được hệ số tương quan giữa 3 biến nêu trên (KL, DL và DN).

```
data WORK.VIDU13
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU13.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat KL best32. ;
informat DL best32. ;
informat DN best32. ;
format KL best12. ;
format DL best12. ;
format DN best12. ;
input
KL
DL
DN
;
PROC CORR;❶
VAR
KL
DL
DN;❷
RUN;
```

Trong đó: ❶ Thủ tục để tính hệ số tương quan (PROC CORR); ❷ Liệt kê các biến để tính hệ số tương quan.

Kết quả từ SAS:

The CORR Procedure❶

3 Variables:KL DL DN

Simple Statistics❷

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
----------	---	------	---------	-----	---------	---------

Simple Statistics ^❷						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
KL	22	61.97727	5.94426	1364	51.50000	71.20000
DL	22	56.75000	2.61551	1249	52.28000	61.15000
DN	22	44.25682	1.34160	973.65000	41.91000	46.27000

Pearson Correlation Coefficients, N = 22 Prob > r under H0: Rho=0 ^❸			
	KL	DL	DN
KL	1.00000	^❹ 0.89667 ^❺ <.0001	0.90527 <.0001
DL	0.89667 <.0001	1.00000	0.64802 0.0011
DN	0.90527 <.0001	0.64802 0.0011	1.00000

Trong đó: ^❶ Kết quả từ câu lệnh PROC CORR để tính hệ số tương quan giữa 3 biến (KL, DL và DN); ^❷ Các thống kê mô tả đối với 3 biến; ^❸ Hệ số tương quan giữa 3 biến; trong 1 ô kết quả; ^❹ giá trị phía trên là hệ số tương quan ($r = 0,89667$); ^❺ Giá trị phía dưới là xác suất đối với hệ số tương quan ($P < 0,0001$).

2.3.2. Hồi quy tuyến tính đơn và đa biến

Có thể xây dựng phương trình hồi quy đơn biến $y = a + bx$ hoặc đa biến $y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$. Với ví dụ sau, ta có thể xây dựng phương trình hồi quy tuyến tính đơn biến quy ước tính khối lượng trứng thông qua đường kính lớn/đường kính bé hoặc đa biến thông qua đường kính lớn và đường kính bé.

Ví dụ 2.13 được sử dụng lại để xây dựng phương trình hồi quy tuyến tính. Với SAS CODE sau đây ta có thể xây dựng phương trình hồi quy tuyến tính với biến phụ thuộc Y là khối lượng trứng và biến độc lập X là đường kính lớn.

Mô hình phân tích:

$$Y = a + bX$$

Trong đó: Y: Biến phụ thuộc - khối lượng trứng (g)

a: tung độ gốc (điểm chắn)

b: hệ số góc đối với biến độc lập

X: Biến độc lập - đường kính lớn (mm)

SAS code của VIDU13A.SAS:

```
data WORK.VIDU13
;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU13.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat KL best32. ;
  informat DL best32. ;
  informat DN best32. ;
  format KL best12. ;
  format DL best12. ;
  format DN best12. ;
  input
    KL
    DL
    DN
  ;
PROC REG;❶
MODEL KL = DL;❷
RUN;
```

Trong đó: ❶ Thủ tục PROC REG để phân tích hồi quy tuyến tính; ❷ Mô hình hồi quy với biến phụ thuộc là khối lượng (KL) và biến độc lập là đường kính lớn (DL).

Kết quả từ SAS:

The REG Procedure❶

Model: MODEL1

Dependent Variable: KL❷

Analysis of Variance❸					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	596.59551	596.59551	82.05	<.0001
Error	20	145.42313	7.27116		
Corrected Total	21	742.01864			

Root MSE	2.69651	R-Square❹	0.8040
Dependent Mean	61.97727	Adj R-Sq❺	0.7942
Coeff Var	4.35080		

Parameter Estimates ^⑨					
Variable	DF	Parameter Estimate ^⑥	Standard Error ^⑦	t Value	Pr > t ^⑧
Intercept	1	-53.67124	12.78032	-4.20	0.0004
DL	1	2.03786	0.22498	9.06	<.0001

Trong đó: **①** Kết quả phân tích hồi quy với **②** Biến phụ thuộc khối lượng (KL); **③** Kết quả phân tích phương sai đối với mô hình, với $P < 0,0001$ mô hình có ý nghĩa; **④** Hệ số xác định R^2 của mô hình (R-Square) = 0,8040 và **⑤** Hệ số xác định hiệu chỉnh $Adj\ R-Sq = \left(1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}\right) = \left(1 - \frac{(1 - 0,8040)(22 - 1)}{22 - 1 - 1}\right) = 0,7942$; **⑥** Các tham số ước tính và **⑦** sai số tiêu chuẩn tương ứng đối với từng hệ số của phương trình hồi quy; **⑧** Xác suất của tương ứng với từng hệ số.

Phương trình hồi quy với biến phụ thuộc là khối lượng (Y) và biến độc lập là đường kính lớn (X) từ phần mềm SAS thu được như sau:

Khối lượng trứng = $-53,67124 + 2,03786 \cdot \text{đường kính lớn}$

⑨ Phần Parameter Estimates, ngoài các hệ số của phương trình hồi quy ta còn có kết quả kiểm định đối với từng hệ số. Với xác suất $P = 0,0004$ và $P < 0,0001$ ta có thể kết luận các hệ số trong phương trình hồi quy khác 0 ($P < 0,05$). Hệ số xác định của phương trình $R^2 = 80,40\%$, hiệu chỉnh $R^2 = 79,4\%$.

Ta cũng có thể xây dựng hồi quy để ước tính khối lượng trứng thông qua 2 biến độc lập là đường kính lớn và đường kính bé với mô hình phân tích sau:

$$Y = a + b_1X_1 + b_2X_2$$

Trong đó: Y: Biến phụ thuộc - khối lượng trứng (g)

A: tung độ gốc

b_1 : hệ số góc đối với biến độc lập thứ nhất - đường kính lớn

b_2 : hệ số góc đối với biến độc lập thứ hai - đường kính bé

X_1 : Biến độc lập thứ nhất - đường kính lớn (mm)

X_2 : Biến độc lập thứ hai - đường kính bé (mm)

Bạn đọc có thể sử dụng lại dữ liệu ở ví dụ 2.13 với thay đổi code đối với mô hình, cụ thể như sau:

SAS CODE VIDU13B.SAS:

```
PROC REG;
```

```
MODEL KL = DL DN; ❶
RUN;
```

Trong đó: ❶ Trong mô hình này, có 2 biến độc lập đó là đường kính lớn (DL) và đường kính bé (DN).

Kết quả từ SAS:

The REG Procedure

Model: MODEL1

Dependent Variable: KL

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	731.05126	365.52563	633.24	<.0001
Error	19	10.96738	0.57723		
Corrected Total	21	742.01864			

Root MSE	0.75976	R-Square	0.9852 ❶
Dependent Mean	61.97727	Adj R-Sq	0.9837 ❷
Coeff Var	1.22586		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-116.55512	5.47204	-21.30	<.0001
DL	1	1.21473	0.08323	14.60	<.0001
DN	1	2.47638	0.16226	15.26	<.0001

Ta có kết quả hoàn toàn tương tự như việc xây dựng phương trình hồi quy đơn biến:

$$Y = -116,55512 + 1,21473X_1 + 2,47638X_2$$

Trong đó: Y: khối lượng trứng,

X_1 : đường kính lớn,

X_2 : đường kính bé.

Điều khác biệt trong trường hợp này là ❶ hệ số xác định $R^2 = 98,5\%$, lớn hơn so với trường hợp đơn biến $R^2 = 80,4\%$; cũng tương tự đối với ❷ hệ số xác định hiệu chỉnh $98,37\%$ so với $79,42\%$.

2.3.3. Lựa chọn mô hình hồi quy tốt nhất

Với trường hợp phương trình hồi quy được xây dựng từ nhiều biến độc lập, cần phải xác định đóng góp của từng biến cho mô hình. Có những biến độc lập khi đưa vào phương trình đóng góp không đáng kể vào mô hình vì vậy cần loại bỏ khi xây dựng phương trình. Để có thể tìm được phương trình hồi quy tốt nhất cần có thêm các lệnh phụ khác.

a. Phương pháp Best Subset

Đây là phương pháp liệt kê tất cả các trường hợp (subset) có thể của các biến độc lập trong mô hình thống kê kèm theo các thông tin (SSE CP AIC) để lựa chọn subset tốt nhất.

SAS CODE VIDU13C.SAS:

```
PROC REG;
MODEL KL = DL DN /SSE CP AIC SELECTION = CP; ❶
RUN;
```

Trong đó: ❶ Lựa chọn (option) đứng ngay sau dấu /SSE CP AIC SELECTION = CP được dùng để chọn subset tốt nhất; trong đó mô hình tốt khi có SSE: tổng bình phương của sai số ngẫu nhiên (phần dư) nhỏ; CP: Conceptual predictive criterion (còn gọi là hệ số Mallows) được sử dụng để xác định mô hình có khả năng giải thích sự biến động lớn nhất với số biến ít nhất có thể và được tính theo công thức $C_p = p + \frac{(MSE - \sigma^2)(n - p)}{\sigma^2}$, trong

đó, MSE: trung bình bình phương sai số ngẫu nhiên của mô hình rút gọn, σ^2 : phương sai của mô hình đầy đủ, n: dung lượng mẫu, p: tham số của mô hình rút gọn, mô hình tốt nhất khi có CP nhỏ và gần bằng p; AIC: Akaike information criterion tính theo công thức $AIC = n \log(SSE/n) + 2p$, mô hình có hệ số AIC bé nhất được coi là mô hình tốt nhất; SELECTION = CP: sắp xếp giá trị CP theo thứ tự tăng dần.

Ngoài các option SSE, Cp, AIC trong các chương trình thống kê còn nhiều option khác như S, PC, SBC nhưng các option khác nhau không phải lúc nào cũng dẫn tới cùng một subset.

Kết quả từ SAS:

The REG Procedure ❶					
Model: MODEL1					
Dependent Variable: KL					
C(p) Selection Method					
Number in Model ❷	C(p) ❸	R-Square ❹	AIC ❺	SSE ❻	Variables in Model ❼

Number in Model ②	C(p) ③	R-Square ④	AIC ⑤	SSE ⑥	Variables in Model ⑦
2	3.0000	0.9852	-9.3146	10.96738	DL DN
1	214.0225	0.8195	43.7381	133.93043	DN
1	233.9326	0.8040	45.5493	145.42313	DL

Trong đó: **①** Kết quả phân tích hồi quy, phương pháp chọn theo hệ số C(p); **②** Số biến độc lập trong mô hình hồi quy; **③** Hệ số C(p), mô hình tối có hệ số C(p) thấp nhất là mô hình tốt nhất nhưng không phải là mô hình tối ưu nhất; **④** Hệ số xác định R²; **⑤** Hệ số AIC; **⑥** Giá trị tổng bình phương của sai số ngẫu nhiên SSE; **⑦** Tên các biến độc lập trong mô hình.

b. Phương pháp Stepwise

- Forward Selection (FORWARD) là phương pháp mà các biến độc lập lần lượt được đưa vào mô hình, mỗi lần đưa một biến vào mô hình gọi là một bước (step). Biến độc lập có giá trị xác suất $P < \alpha$ sẽ được đưa vào trong mô hình.

SAS CODE VIDU13D.SAS

```
data WORK.VIDU13
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU13.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat KL best32. ;
informat DL best32. ;
informat DN best32. ;
format KL best12. ;
format DL best12. ;
format DN best12. ;
input
KL
DL
DN
;
PROC REG; ①
MODEL KL = DL DN / SLENTY=0.15 SELECTION=FORWARD; ②
RUN;
```

Trong đó: **①** Thủ tục PROC REG được sử dụng; **②** Mô hình phân tích hồi quy bao gồm 2 biến độc lập (DL DN) và 1 biến phụ thuộc (KL), SLENTY=0.15 (Significance Level of Entry = 0.15) tức biến có giá trị xác suất $P < 0,15$ mới có thể thêm vào mô hình hồi quy.

*The REG Procedure **①***

Model: MODEL1

Dependent Variable: KL ②

Forward Selection: Step 1 ③

Variable DN Entered: R-Square = 0.8195 and C(p) = 214.0225 ④

Analysis of Variance ⑤					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	608.08820	608.08820	90.81	<.0001
Error	20	133.93043	6.69652		
Corrected Total	21	742.01864			

Variable	⑥ Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-115.53651	18.63648	257.37130	38.43	<.0001
DN	4.01099	0.42091	608.08820	90.81	<.0001

Bounds on condition number: 1, 1

Forward Selection: Step 2 ⑦

Variable DL Entered: R-Square = 0.9852 and C(p) = 3.0000 ⑧

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	731.05126	365.52563	633.24	<.0001
Error	19	10.96738	0.57723		
Corrected Total	21	742.01864			

Variable	⑨ Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-116.55512	5.47204	261.88686	453.70	<.0001
DL	1.21473	0.08323	122.96306	213.02	<.0001
DN	2.47638	0.16226	134.45575	232.93	<.0001

Bounds on condition number: 1.7239, 6.8957

All variables have been entered into the model.

Summary of Forward Selection ⑩							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F

Summary of Forward Selection ¹⁰							
Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	DN	1	0.8195	0.8195	214.023	90.81	<.0001
2	DL	2	0.1657	0.9852	3.0000	213.02	<.0001

Trong đó: ❶ Kết quả phân tích hồi quy với PROC REG; ❷ đối với biến khối lượng (KL); ❸ Bước 1 (Step 1); ❹ đưa biến độc lập thứ nhất (DN) vào trong mô hình với thông tin thu được $R^2 = 0,8195$ và $C(p) = 214,0225$; ❺ Kết quả phân tích phương sai của mô hình; ❻ Các tham số ước tính ở Step 1 (tọa độ gốc -115,53651 và hệ số góc đối với biến độc lập DN = 4,01099); ❼ Bước 2 (Step 2); ❽ đưa thêm biến độc lập thứ hai (DL) vào trong mô hình với thông tin thu được $R^2 = 0,9852$ và $C(p) = 3,0000$; ❾ Các tham số ước tính ở Step 2 (tọa độ gốc -116,55512 và hệ số góc đối với biến độc lập DN = 2,47638 và DL = 1,21473), tất cả các biến độc lập (DL, DN) đều có trong mô hình cuối cùng; ❿ Các thông tin tổng hợp cho phần Forward được thể hiện chi tiết trong bảng.

- Backward Elimination (BACKWARD) là phương pháp loại bỏ dần những biến độc lập trong phương trình hồi quy (ngược lại với phương pháp Forward Selection). Biến độc lập có giá trị xác suất $P > \alpha$ sẽ bị loại bỏ khỏi mô hình.

SAS CODE VIDU13E.SAS

```
data WORK.VIDU13          ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU13.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat KL best32. ;
  informat DL best32. ;
  informat DN best32. ;
  format KL best12. ;
  format DL best12. ;
  format DN best12. ;
input
  KL
  DL
  DN
;
PROC REG; ❶
```

```
MODEL KL = DL DN / SLSTAY=0.15 SELECTION=BACKWARD; ❷
RUN;
```

Trong đó: ❶ Thủ tục PROC REG được sử dụng; ❷ Mô hình phân tích hồi quy bao gồm 2 biến độc lập (DL DN) và 1 biến phụ thuộc (KL), SLSTAY=0.15 (Significance Level of Stay = 0,15) tức biến có giá trị xác suất $P \leq 0,15$ sẽ có mặt (stay) trong mô hình hồi quy và biến có $P > 0,15$ sẽ bị loại (elimination) khỏi mô hình.

The REG Procedure ❶

Model: MODEL1

Dependent Variable: KL ❷

Backward Elimination: Step 0 ❸

All Variables Entered: R-Square = 0.9852 and C(p) = 3.0000 ❹

Analysis of Variance ❺					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	731.05126	365.52563	633.24	<.0001
Error	19	10.96738	0.57723		
Corrected Total	21	742.01864			

Variable	❻ Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-116.55512	5.47204	261.88686	453.70	<.0001
DL	1.21473	0.08323	122.96306	213.02	<.0001
DN	2.47638	0.16226	134.45575	232.93	<.0001

Bounds on condition number: 1.7239, 6.8957

All variables left in the model are significant at the 0.1500 level. ❼

Trong đó: ❶ Kết quả phân tích hồi quy với PROC REG; ❷ đối với biến khối lượng (*KL*); ❸ Bước 0 (Step 0); ❹ đưa tất cả các biến độc lập vào trong mô hình với thông tin thu được $R^2 = 0,9852$ và $C(p) = 3,0000$; ❺ Kết quả phân tích phương sai của mô hình; ❻ Các tham số ước tính ở Step 0 (tọa độ gốc -116,55512 và hệ số góc đối với biến độc lập $DL = 1,21473$ và $DN = 2,47638$); ❼ Không biến nào bị loại khỏi mô hình (tất cả các biến độc lập (DL, DN) đều có trong mô hình cuối cùng ở mức $P = 0,15$).

Phương pháp Stepwise chỉ dựa trên phân tích phần dư trong phân tích hồi quy bội tuyến tính để lựa chọn biến đưa vào và biến đưa ra. Để đảm bảo trong mô hình cuối cùng có mặt một số biến có ý nghĩa về mặt chuyên môn các nhà thống kê thường đưa thêm phần chọn biến bắt buộc có mặt trước khi bắt đầu việc lọc theo Stepwise.

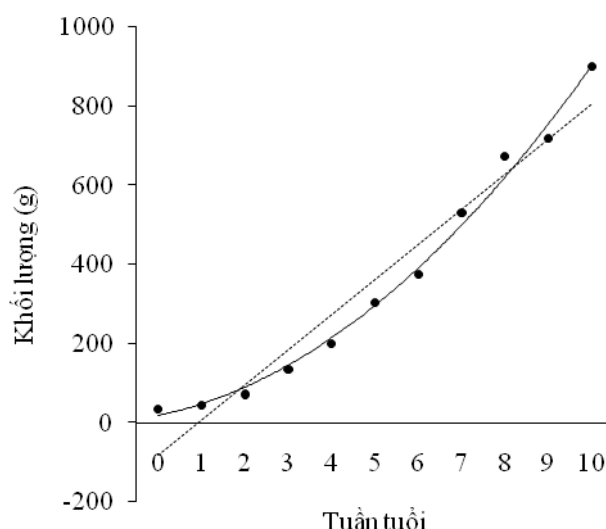
2.3.4. Hồi quy đa thức

Ở mục 2.3.1, mối liên hệ tuyến tính giữa các biến đã được đề cập đến. Trong sinh học mối liên hệ giữa các biến có thể tồn tại nhưng ở dưới dạng phi tuyến. Trong những trường hợp như vậy mô hình tuyến tính nêu trên sẽ không còn phù hợp.

Ví dụ 2.14: Khối lượng của gà Hồ (gram) qua các tuần tuổi (từ 0 đến 10 tuần tuổi) được trình bày ở bảng sau:

Tuần tuổi (tuần)	0	1	2	3	4	5	6	7	8	9	10
Khối lượng (gram)	35	45	72	134	199	303	374	529	672	717	899

Nếu thể hiện mối liên hệ giữa khối lượng và tuần tuổi bằng các điểm (•) trên đồ thị ta thấy mối liên hệ giữa 2 biến này không phải là tuyến tính (xem đường ---) mà có mối liên hệ phi tuyến bậc 2 (xem đường —).



Ta có mô hình phân tích: $y = a + b_1x + b_2x^2$

Số liệu có thể nhập vào file VIDU14.XLS định dạng excel sau đó chuyển file VIDU14.TXT định dạng txt với 2 cột lần lượt là khối lượng (KL) và tuần tuổi (TUAN).

SAS code của VIDU14.SAS:

```
data WORK.VIDU14 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU14.txt' delimiter='09'x MISSOVER DSD lrecl=32767
  firstobs=2 ;
  informat KL best32. ;
  informat TUAN best32. ;
  format KL best12. ;
  format TUAN best12. ;
  input ❶
        KL
```

```

TUAN
;
TUAN2 = TUAN*TUAN; ❷
PROC REG; ❸
MODEL KL = TUAN TUAN2; ❹
RUN;

```

Trong đó: ❶ Trong file số liệu có 2 cột số liệu (KL và TUAN); ❷ $TUAN2 = TUAN \times TUAN$; câu lệnh sử dụng để tạo ra cột mới (TUAN2) bằng bình phương của cột tuần (TUAN*TUAN); như vậy trong bộ số liệu sẽ có 3 cột, bao gồm: 2 cột đã có sẵn (khối lượng và tuần) và cột thứ 3 (TUAN2) bằng bình phương của cột tuần; ❸ Câu lệnh chạy mô hình hồi quy; ❹ Mô hình hồi quy, trong đó khối lượng (KL) là biến phụ thuộc, biến tuần (TUAN) và tuần bình phương (TUAN2) là biến độc lập trong mô hình.

Kết quả từ SAS:

The REG Procedure ❶
Model: MODEL1
Dependent Variable: KL ❷

Analysis of Variance ❸					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	905796	452898	576.73	<.0001
Error	8	6282.31329	785.28916		
Corrected Total	10	912078			

Root MSE	28.02301	R-Square	0.9931 ❹
Dependent Mean	361.72727	Adj R-Sq	0.9914 ❺
Coeff Var	7.74700		

Parameter Estimates ❻					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t ❿
Intercept ❼	1	16.67832	21.34941	0.78	0.4572
TUAN ❽	1	22.72657	9.93300	2.29	0.0514
TUAN2 ❾	1	6.61189	0.95669	6.91	0.0001

Trong đó: ❶ Kết quả phân tích hồi quy; ❷ Đối với biến phụ thuộc khối lượng (KL); ❸ Bảng kết quả phân tích phương sai; ❹ Hệ số xác định của mô hình (R-Square) = 0,9931, nghĩa là 99,31% biến động được giải thích bằng mô hình hồi quy;

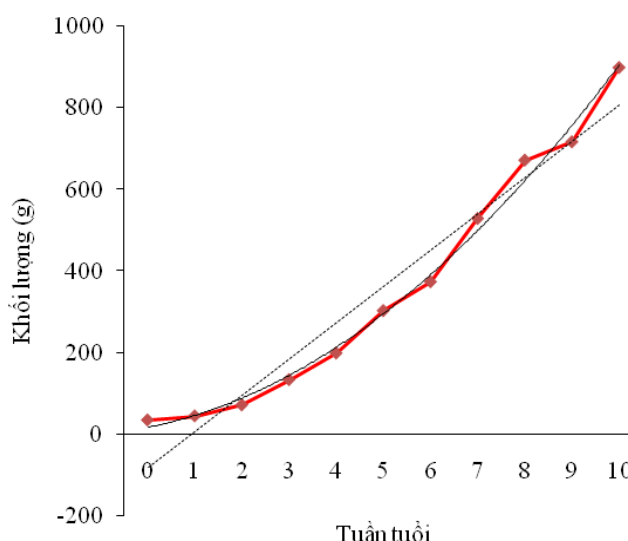
⑤ Hệ số xác định hiệu chỉnh theo các biến độc lập trong mô hình (Adj R-Sq) = 0,9914.

$$\text{Adj R-Sq} = \left(1 - \frac{(1 - R^2)(N - 1)}{N - p - 1} \right)$$
, trong đó R^2 , N và p lần lượt là hệ số xác định, dung

lượng mẫu và số biến độc lập; ⑥ Các hệ số ước tính bao gồm ⑦ tung độ gốc (Intercept) = 16,67832, ⑧ hệ số góc đối với biến thứ nhất (TUAN) = 22,72657 và ⑨ hệ số góc đối với biến thứ 2 (TUAN2); ⑩ Xác suất đối với từng hệ số.

Dựa trên kết quả phân tích, mô hình hồi quy có thể được viết như sau:

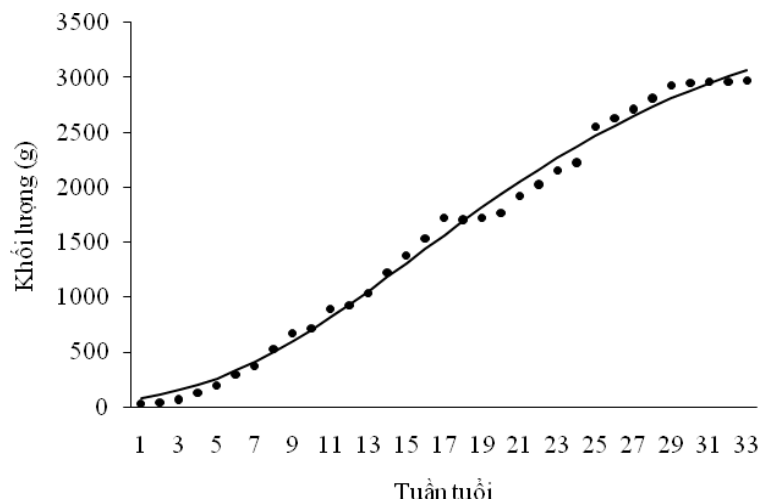
$$\text{Khối lượng} = 16,67832 + 22,72657 * \text{Tuần} + 6,61189 * \text{Tuần}^2$$



2.3.5. Hồi quy phi tuyến

Ví dụ 2.15: Khối lượng của gà Hồ (gram) qua các tuần tuổi (từ 0 đến 32 tuần tuổi) được trình bày ở bảng sau:

Tuần tuổi	0	1	2	3	4	5	6	7	8	9	10
Khối lượng	35	45	72	134	199	303	374	529	672	717	899
Tuần tuổi	11	12	13	14	15	16	17	18	19	20	21
Khối lượng	933	1042	1228	1386	1541	1729	1711	1726	1771	1927	2032
Tuần tuổi	22	23	24	25	26	27	28	29	30	31	32
Khối lượng	2158	2232	2557	2632	2718	2818	2937	2960	2964	2967	2977



Số liệu có thể phân tích theo hàm sinh trưởng Gompertz theo mô hình sau:

$$Y = me^{-ae^{-bx}} = m * EXP[-a * EXP(-bx)]$$

Trong đó: Y: khối lượng gà (g),

m: khối lượng tiệm cận trên – khối lượng trưởng thành (g),

a: hằng số tích hợp liên quan đến khối lượng sơ sinh (0 tuần tuổi),

b: tỷ lệ tốc độ tăng trưởng tối đa so với khối lượng trưởng thành của gà,

x: tuổi (tuần),

EXP: cơ số logarit tự nhiên e (2,71828).

Số liệu có thể nhập vào file VIDU15.XLS định dạng excel sau đó chuyển file VIDU15.TXT định dạng txt với 2 cột lần lượt là tuần tuổi (TUAN) và khối lượng của gà Hồ (KL).

SAS CODE của VIDU15.SAS:

```
data WORK.VIDU15          ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU15.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
    informat TUAN best32. ;
    informat KL best32. ;
    format TUAN best12. ;
    format KL best12. ;
```

```

input
    TUAN
    KL
;

PROC NLIN; ❶

PARMS m = 2980 a = 4.4 b = 0.14 ; ❷

MODEL KL = m*exp(-a*exp(-b*TUAN)) ; ❸

RUN;

```

Trong đó: ❶ Thủ tục PROC NLIN để xây dựng mô hình hồi quy phi tuyến; ❷ Các tham số sử dụng làm giá trị «Starting Value»; các giá trị này đưa vào mô hình dựa trên số liệu của mô hình và/hoặc các nghiên cứu tương tự. Ví dụ giá trị $m = 2980$ dựa vào khối lượng của gà Hồ ở tuần tuổi 32 (2977g), giá trị a và b lấy từ kết quả nghiên cứu của Moula et al. (2011); ❸ Mô hình Gompertz nhằm xây dựng phương trình hồi quy phi tuyến ước tính khối lượng gà (KL) thông qua tuần tuổi (TUAN).

Kết quả từ SAS:

The NLIN Procedure ❶

Dependent Variable KL

Method: Gauss-Newton

Iterative Phase ❷				
Iter	m	a	b	Sum of Squares
0	2980.0	4.4000	0.1400	1268948
1	3187.6	3.8926	0.1128	506210
2	3582.9	3.7918	0.0930	374287
3	3728.2	3.8696	0.0934	239655
4	3734.6	3.8619	0.0932	239613
5	3734.3	3.8626	0.0932	239613
6	3734.3	3.8625	0.0932	239613

NOTE: Convergence criterion met.

NOTE: An intercept was not specified for this model.

❸ Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
----------	----	----------------	-------------	---------	---------------

③ Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Model	3	1.1202E8	37339372	4674.96	<.0001
Error	30	239613	7987.1		
Uncorrected Total	33	1.1226E8			

④ Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
m	3734.3	155.3	3417.1	4051.6
a	3.8625	0.1907	3.4732	4.2519
b	0.0932	0.00607	0.0808	0.1056

Approximate Correlation Matrix ⑤			
	m	a	b
m	1.0000000	-0.6384440	-0.9353337
a	-0.6384440	1.0000000	0.8471660
b	-0.9353337	0.8471660	1.0000000

Trong đó: ① Kết phân tích hồi quy bằng thủ tục PROC NLIN đối với biến phụ thuộc khối lượng (KL); ② Kết quả xử lý để tìm được giá trị trung bình bình phương bé nhất (Sum of Squares) sau 6 lượt chạy, hàng đầu tiên (0) là giá trị xuất phát «Starting Value»; ③ Kết quả phân tích mô hình, với $P < 0,0001$ tức mô hình có ý nghĩa thống kê và hệ số xác định của mô hình $R^2 = 99,79\%$; ④ Các tham số ước tính ($m = 3734,3$, $a = 3,8625$ và $b = 0,0932$), sai số tiêu chuẩn (Approx Std Error) và khoảng tin cậy 95% (Approximate 95% Confidence Limits) đối với từng hệ số đối với phương trình hồi quy; ⑤ Ma trận tương quan giữa các hệ số của phương trình hồi quy.

Dựa trên kết quả phân tích, mô hình hồi quy phi tuyến để ước tính khối lượng gà Hồ như sau:

$$\text{Khối lượng} = 3734,3 * \text{EXP} (-3,8625 * \text{EXP} (-0,0932 * \text{tuần}))$$

Tuổi tại điểm uốn:

$$PI = \ln(a)/b = \ln(3,8625)/0,0932 = 14,5 \text{ tuần}$$

Khối lượng bò tại điểm uốn:

$$YPI = m/e = 3.734,3/2,71828 = 1.373,77$$

Tăng khối lượng tuyệt đối (g/tuần) tại điểm uốn:

$$MWGPI = mb/e = 3.734,3 * 0,0932 / 2,71828 = 128g.$$

2.4. PHÂN TÍCH SỐ LIỆU ĐỊNH TÍNH

2.4.1. Phép thử χ^2 và Fisher exact test

a. So sánh tỷ lệ

Khi so sánh các tỷ lệ hoặc nghiên cứu mối liên hệ giữa các yếu tố đối với biến định tính ta luôn đặt giả thiết H_0 : Không có sự sai khác có ý nghĩa thống kê giữa các tỷ lệ hoặc không có mối liên hệ giữa các yếu tố (tùy theo mục tiêu của bài toán đặt ra).

Ví dụ 2.17: Một thí nghiệm được tiến hành nhằm đánh giá sự liên hệ giữa tỷ lệ viêm nội mạc tử cung và giống. Trong tổng số 700 bò sữa trong nghiên cứu thuần tập (cohort studies), có 500 con giống Holstein Friesian và 200 con giống Jersey. Kết quả nghiên cứu thu được như sau:

Giống	Viêm nội mạc tử cung		Tổng số
	Có	Không	
Holstein	100	400	500
Jersey	10	190	200
Tổng số	110	590	700

SAS CODE của VIDU17.SAS:

```
DATA SAS217; ❶
INPUT GIONG $ BENH $ SOLUONG; ❷
CARDS; ❸
H      C      100
H      K      400
J      C      10
J      K      190
;
PROC FREQ; ❹
WEIGHT SOLUONG; ❺
TABLE GIONG*BENH / CHISQ EXPECTED; ❻
RUN;
```

Do số liệu định tính thu thập được từ các thí nghiệm có số lượng ít và cấu trúc đơn giản nên bạn đọc có thể chọn cách nhập số liệu trực tiếp qua cửa sổ Editor (xem chi tiết mục a phần 2.1.2 chương 2), trong đó: ❶ Tạo cơ sở dữ liệu SAS217 trong chương trình; ❷ Với 3 cột số liệu lần lượt là *giống* (GIONG), *bệnh* (BENH) và *số lượng*

(SOLUONG); ❸ Câu lệnh bắt đầu nhập số liệu thô; ❹ Thủ tục PROC FREQ để tiến hành phân tích các biến định tính; ❺ Cột có số lượng các ca bệnh/không bệnh; ❻ Sử dụng phép thử χ^2 (CHISQ) để so sánh tỷ lệ mắc bệnh giữa các giống, EXPECTED tính tần suất ước tính.

Kết quả từ SAS:

The FREQ Procedure❶			
Frequency❷ Expected Percent Row Pct Col Pct	Table of GIONG by BENH❸		
	GIONG	BENH	
		C	K
	H	100	400
		78.571	421.43
		14.29	57.14
		20.00	80.00
		90.91	67.80
	J	10	190
		31.429	168.57
		1.43	27.14
		5.00	95.00
		9.09	32.20
	Total	110	590
		15.71	84.29
			700
			100.00

Trong đó, ❶ kết quả phân tích của thủ tục PROC FREQ ❷ giải thích các giá trị tương ứng trong mỗi ô theo thứ tự từ trên xuống dưới (Frequency: tần suất quan sát, Expected Percent: tần suất ước tính, Percent: %, Row Pct: % theo hàng, Col Pct: % theo cột); ❸ Bảng tương liên 2 × 2 theo giống và bệnh. Ví dụ đối với ô thứ nhất các giá trị lần lượt là: 1) tần suất quan sát (100), 2) tần suất ước tính (78,571) 3) phần trăm (14,29) 4) phần trăm theo hàng (20%) và 5) phần trăm theo cột (90,91).

Statistics for Table of GIONG by BENH❶			
Statistic	DF	Value	Prob
Chi-Square	1❷	24.2681❸	<.0001❹
Likelihood Ratio Chi-Square	1	29.0537	<.0001
Continuity Adj. Chi-Square	1	23.1488	<.0001
Mantel-Haenszel Chi-Square	1	24.2334	<.0001
Phi Coefficient		0.1862	
Contingency Coefficient		0.1830	

Statistic	DF	Value	Prob
Cramer's V		0.1862	

Fisher's Exact Test ❸	
Cell (1,1) Frequency (F)	100
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	8.496E-08
Table Probability (P)	6.784E-08
Two-sided Pr <= P	1.371E-07

Sample Size = 700 ❹

Trong đó: ❶ Kết quả phân tích χ^2 ; ❷ bậc tự do $df = 1$; ❸ Giá trị Chi-Square $\chi^2 = 24,2681$; và ❹ xác suất của phép thử $P < 0,0001$. Với xác suất này giả thiết H_0 bị bác bỏ và kết luận *Có mối liên hệ giữa bệnh viêm nội mạc tử và giống bò* ($P < 0,001$).

Lưu ý: Đối với trường hợp mẫu bé (tần suất ước tính < 5), thế thay thế phép thử χ^2 bằng ❺ phép thử Fisher exact test (xem phần b dưới đây) ❻ Dung lượng mẫu của thí nghiệm.

b. Kiểm định sự phù hợp

Phần này tập trung giới thiệu kiểm định định luật Mendel hay Hardy-Weinberg trong di truyền bằng phép thử χ^2 hay phép thử chính xác của Fisher.

Ví dụ 2.17a: Tần số kiểu gen AA, AB và BB của đa hình T3737C gen insulin ở gà Hồ lần lượt là 42, 13 và 0. Tần số kiểu gen của đa hình T3737C gen insulin ở gà Hồ có tuân theo định luật Hardy-Weinberg?

Từ số liệu nêu trên ta có thể ước tính được:

Kiểu gen	AA	AB	BB	Tổng số
Giá trị quan sát	42	13	0	55
Tần suất ước tính	$0,882^2 = 0,778$ (p^2)	$2 * 0,882 * 0,118 = 0,208$ ($2pq$)	$0,118^2 = 0,014$ (q^2)	1
Giá trị ước tính	$55 * 0,88 = 42,77$	$2 * 55 * 0,882 * 0,11 = 11,46$	$55 * 0,118^2 = 0,77$	55

Tần suất ước tính được xác định từ tần số allele A (P_A) và B (P_B).

$$P_A = \left(\frac{42 * 2 + 13}{55 * 2} \right) = 0,882$$

$$P_B = 1 - 0,882 = 0,118$$

$\chi^2 = (42 - 42,77)^2/42,77 + (13 - 11,46)^2/11,46 + (0 - 0,77)^2/0,77 = 0,9973$ với bậc tự do $df = 3 - 1 = 2$. Giá trị xác suất của phép thử $p = 0,6073$ (đối với kiểm định χ^2).

SAS CODE của VIDU17A.SAS:

```
DATA SAS217A; ❶
INPUT KIEUGEN $ TANSO; ❷
CARDS; ❸
AA      42
AB      13
BB       0
;
PROC FREQ; ❹
WEIGHT TANSO / ZEROS; ❺
TABLES KIEUGEN / TESTP = (0.778 0.208 0.014); ❻
EXACT CHISQ; ❼
RUN;
```

Trong đó ❶ Tạo cơ sở dữ liệu SAS217A trong chương trình; ❷ Với 2 cột số liệu lần lượt là *kiểu gen* (KIEUGEN) và *tần số* (TANSO); ❸ Câu lệnh bắt đầu nhập số liệu thô; ❹ Thủ tục PROC FREQ để tiến hành phân tích các biến định tính; ❺ Cột có số lượng tần số của các kiểu gen tương ứng, / ZEROS sử dụng số liệu bằng không để tính; ❻ Kiểm định tần số kiểu gen thực tế với tần số kiểu gen lý thuyết theo định luật Hardy-Weinberg với tần suất của các kiểu gen AA, AB và BB tương ứng là 0,778; 0,208 và 0,014 ❼ Sử dụng kiểm định chính xác của Fisher vì có một giá trị có ước tính bé hơn 5 (xem phần kết quả ước tính từ ví dụ 2.17a).

Kết quả từ SAS:

The FREQ Procedure ❶

KIEUGEN ❷	Frequency ❸	Percent ❹	Test ❺ Percent	Cumulative Frequency	Cumulative Percent
AA	42	76.36	77.80	42	76.36
AB	13	23.64	20.80	55	100.00
BB	0	0.00	1.40	55	100.00

Chi-Square Test ❻
for Specified Proportions

Chi-Square	0.9973 ❼
------------	----------

Chi-Square Test ^⑥ for Specified Proportions	
DF	2 ^⑧
Asymptotic Pr > ChiSq	0.6073 ^⑨
Exact Pr >= ChiSq	0.6354 ^⑩
WARNING: 33% of the cells have expected counts less than 5. (Asymptotic) Chi-Square may not be a valid test.	

Sample Size = 55

① Kết quả phân tích của thủ tục PROC FREQ ; ② Kiểu gen; ③ Tần suất quan sát tương ứng với từng kiểu gen; ④ Phần trăm tương ứng của từng kiểu gen; ⑤ Phần trăm lý thuyết (tỷ lệ cần kiểm định); ⑥ Kết quả phân tích χ^2 ; ⑦ Giá trị Chi-Square $\chi^2 = 0,9973$; ⑧ bậc tự do $df = 2$; ⑨ xác suất bằng phép thử χ^2 là $P = 0,6073$ và ⑩ bằng phép thử chính xác của Fisher là $P = 0,6354$. Vì có một giá trị ước tính bé hơn 5 nên xác suất từ phép thử chính xác của Fisher ($P = 0,6354$) sẽ được dùng trong trường hợp này. Kết luận: Tần số kiểu gen của đa hình T3737C gen insulin ở gà Hồ tuân theo định luật Hardy-Weinberg ($P = 0,6354$).

2.4.2. Hồi quy logistic

Hồi quy logistic có thể sử dụng để đánh giá ảnh hưởng của yếu tố nghiên cứu đến tỷ lệ mắc bệnh. Trong trường này biến phụ thuộc là biến định tính và biến độc lập là biến liên tục. Ví dụ dưới đây minh họa việc sử dụng mô hình hồi quy logit để phân tích số liệu định tính.

Ví dụ 2.18: Một thí nghiệm nhằm nghiên cứu ảnh hưởng của tuổi đẻ lần đầu (tháng) đến tỷ lệ viêm vú ở bò. Số liệu thu thập trên 21 bò về tình trạng viêm vú (0: không mắc bệnh và 1: mắc bệnh) như sau:

Bò số	1	2	3	4	5	6	7	8	9	10	11
Tuổi (tháng)	19	20	20	20	21	21	21	22	22	22	23
Viêm vú	1	1	0	1	0	1	1	1	1	0	1
Bò số	12	13	14	15	16	17	18	19	20	21	
Tuổi (tháng)	26	27	27	27	27	29	30	30	31	32	
Viêm vú	1	0	1	0	0	1	0	0	0	0	

Số liệu có thể phân tích theo phương trình hồi quy logistic sau:

$$\log[p_i / (1-p_i)] = \beta_0 + \beta_1 x_i$$

Trong đó: p_i : tỷ lệ mắc bệnh của bò thứ i ,

x_i : tuổi đẻ lần đầu của bò thứ i ,

β_0 : hệ số tung độ gốc,

β_1 : hệ số góc.

Số liệu có thể nhập vào file VIDU18.XLS định dạng excel sau đó chuyển file VIDU18.TXT định dạng txt với 2 cột lần lượt là tuần tuổi (TUAN) và khối lượng của gà Hồ (KL).

SAS code của VIDU18.SAS:

```
data WORK.VIDU18 ;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU18.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat TUOI best32. ;
informat VIEMVU best32. ;
format TUOI best12. ;
format VIEMVU best12. ;
input
    TUOI
    VIEMVU
;
PROC GENMOD DESC; ❶
MODEL VIEMVU = TUOI /DIST = BIN LINK = LOGIT TYPE3; ❷
RUN;
```

Trong đó: ❶ Thủ tục PROCEDURE GENMOD được sử dụng để phân tích hồi quy logit, DESC để chọn giá trị 1 đối với trường hợp mắc bệnh ; ❷ Mô hình phân tích ảnh hưởng của tuổi đẻ lần đầu (TUOI) đến bệnh viêm vú (VIEMVU), DIST = BIN phân bố nhị thức (0 hoặc 1 tương ứng mắc bệnh hoặc không mắc bệnh), LINK = LOGIT mô hình sử dụng là mô hình logit, TYPE3 tính trực tiếp xác suất cho yếu tố tuổi (TUOI).

Kết quả từ SAS:

The GENMOD Procedure ❶	
Model Information ❷	
Data Set	WORK.MATITIS
Distribution	Binomial

Model Information ²	
Link Function	Logit
Dependent Variable	VIEMVU
Observations Used	21

Response Profile ³		
Ordered Value	VIEMVU	Total Frequency
1	1	11
2	0	10

PROC GENMOD is modeling the probability that VIEMVU='1'.

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	19	23.8416	1.2548
Scaled Deviance	19	23.8416	1.2548
Pearson Chi-Square	19	20.4851	1.0782
Scaled Pearson X2	19	20.4851	1.0782
Log Likelihood		-11.9208	

Algorithm converged.

Analysis Of Parameter Estimates ⁴							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	6.7439	3.2640	0.3466	13.1412	4.27	0.0388
TUOI	1	-0.2701	0.1315	-0.5278	-0.0124	4.22	0.0399
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis			
Source	DF	Chi-Square	Pr > ChiSq
TUOI	1	5.22	0.0223 ⁵

Trong đó; ❶ Kết phân tích từ PROCEDURE GENMODE; ❷ Thông tin về mô hình phân tích; và ❸ cấu trúc số liệu; ❹ Hệ số hồi quy $\beta_0 = 6,7439$ và $\beta_1 = -0,2701$; ❺ Xác suất $P = 0,0223$ nên kết luận tuổi đẻ lứa đầu có ảnh hưởng đến tỷ lệ mắc bệnh viêm vú ở bò.

2.5. BIẾN ĐỔI SỐ LIỆU

Khi số liệu có phân bố chuẩn và các phương sai đồng nhất, việc phân tích số liệu và đưa ra các kết luận sẽ đơn giản. Trong một số trường hợp các điều kiện nêu trên không được thỏa mãn bạn đọc sẽ có 2 lựa chọn: (1) biến đổi số liệu và (2) sử dụng thống kê phi tham số (xem mục 2.6). Mục này sẽ đề cập đến cách biến đổi số liệu để có thể thỏa mãn điều kiện của bài toán.

Có 3 cách biến đổi dữ liệu thường được sử dụng đối với các biến (x) trong sinh học, bao gồm:

(1) logarit: $x' = \log(x)$ sử dụng trong trường hợp số liệu có xu hướng lệch phải. Logarit cơ số 10 và cơ số e được sử dụng phổ biến nhất sẽ sử dụng hàm logarit.

(2) bình phương: $x' = x^2$ sử dụng trong trường hợp số liệu lệch trái.

(3) logistics: $x' = \log[x/(1-x)]$ sử dụng trong trường hợp tỷ lệ/phần trăm đặc biệt là với hàm phân bố nhị thức (xem ví dụ 2.18 phần 2.4.2).

Ngoài ra còn có cách biến đổi sử dụng arcsin hay biến đổi tổng quát BOX COX.

Xem xét ví dụ sau đây:

Ví dụ 2.19. Thí nghiệm được tiến hành nhằm đánh giá ảnh hưởng của chế phẩm sinh học (A, B và C) đến lượng vi sinh vật trong phân chuồng. Thí nghiệm được tiến hành qua 2 mùa (hè và đông). Số liệu thu được trình bày ở bảng dưới đây. Tiến hành phân tích số liệu và cho biết ảnh hưởng của yếu tố thí nghiệm. Lưu ý kiểm tra phân bố chuẩn của số liệu.

Chế phẩm	Hè	Đông
A	3.800.000	13.000.000
A	1.000.000	1.600.000
A	320.000	360.000
B	1.600.000	400.000
B	100.000	1.100.000
B	100.000	1.000.000
C	22.000.000	1.600.000
C	11.000.000	1.800.000
C	14.000.000	120.000.000

Mô hình phân tích:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Trong đó: y_{ijk} : quan sát thứ k ở chế phẩm thứ i và mùa j

μ : trung bình chung

α_i : ảnh hưởng của chế phẩm i,

β_j : ảnh hưởng mức của mùa j,

$(\alpha\beta)_{ij}$: tương tác giữa mức i của chế phẩm và mức j của mùa

ε_{ijk} : sai số ngẫu nhiên, độc lập, phân phối chuẩn $N(0, \sigma^2)$

Số liệu có thể nhập vào file VIDU19.XLS định dạng excel sau đó chuyển file VIDU19.TXT định dạng txt với 3 cột lần lượt là chế phẩm (CP), mùa (MUA) và số lượng vi sinh vật (VSV).

Trước khi biến đổi số liệu:

SAS CODE của VIDU19.SAS:

```
data WORK.VIDU19 ;
  %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
  infile 'D:\SAS2014\VIDU19.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
  informat CP best32. ;
  informat MUA best32. ;
  informat VSV comma32. ;
  format CP best12. ;
  format MUA best12. ;
  format VSV comma12. ;
  input
    CP
    MUA
    VSV
  ;
PROC GLM; ❶
CLASS CP MUA;
MODEL VSV = CP MUA CP*MUA / SS4; ❷
OUTPUT OUT = DLMOI R = PHANDU; ❸
RUN;

PROC UNIVARIATE NORMAL DATA = DLMOI; ❹
VAR PHANDU; ❺
RUN;
```

Trong đó: ❶ Câu lệnh PROC GLM được sử dụng để phân tích số liệu; ❷ Mô hình phân tích các yếu tố ảnh hưởng (CP MUA CP*MUA) đến số lượng vi sinh vật (VSV); ❸ Câu lệnh để tạo ra cơ sở dữ liệu (DLMOI) với cột mới phần dư (PHANDU) chính là phần

ε trong mô hình phân tích; ❹ Kiểm định phân bố chuẩn của phần dư nhằm kiểm tra điều kiện của bài toán (số liệu phân bố chuẩn và phương sai đồng nhất) với ❺ cột số liệu phần dư (PHANDU).

Kết quả từ SAS:

The GLM Procedure ❶

Class Level Information		
Class	Levels	Values
CP	3	1 2 3
MUA	2	1 2

Number of observations	18
------------------------	----

The GLM Procedure

Dependent Variable: VSV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3.7909526E15	7.5819053E14	0.96	0.4803
Error	12	9.5002933E15	7.9169111E14		
Corrected Total	17	1.3291246E16			

R-Square	Coeff Var	Root MSE	VSV Mean
0.285222	260.0196	28137006	10821111

Source ❷	DF	Type IV SS	Mean Square	F Value	Pr > F
CP	2	2.8019067E15	1.4009534E15	1.77	0.2121
MUA	1	4.199202E14	4.199202E14	0.53	0.4804
CP*MUA	2	5.6912573E14	2.8456287E14	0.36	0.7053

Trong đó: ❶ Kết quả phân tích của thủ tục GLM; ❷ Kết quả phân tích phương sai đối với các yếu tố trong mô hình. Kết quả cho thấy không ảnh hưởng của chế phẩm ($P = 0,2121$), mùa ($P = 0,4804$) và không có tương tác giữa 2 yếu tố này ($P = 0,7053$) đến số lượng vi sinh vật.

The UNIVARIATE Procedure ❶

Variable: PHANDU

Moments			
N	18	Sum Weights	18
Mean	0	Sum Observations	0
Std Deviation	23639813.5	Variance	5.58841E14
Skewness	1.84559801	Kurtosis	8.07135645
Uncorrected SS	9.50029E15	Corrected SS	9.50029E15
Coeff Variation	.	Std Error Mean	5571957.49

Basic Statistical Measures			
Location		Variability	
Mean	0	Std Deviation	23639814
Median	-500000	Variance	5.58841E14
Mode	-500000	Range	118400000
		Interquartile Range	4386667

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > t	1.0000
Sign	M	-2	Pr >= M	0.4807
Signed Rank	S	-18.5	Pr >= S	0.4359

Tests for Normality ²				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.648669	Pr < W	<0.0001 ³
Kolmogorov-Smirnov	D	0.311758	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.560177	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	2.780733	Pr > A-Sq	<0.0050

Kiểm tra phân bố chuẩn của cột **PHANDU** ❷ Kết quả kiểm tra phân bố chuẩn, sử dụng phép thử Shapiro-Wilk ($n = 36 < 2.000$) ❸ giá trị $P < 0,0001$, như vậy H_0 bị bác bỏ và chấp nhận H_1 tức là cột phần dư không có phân bố chuẩn. Kết luận: Số liệu không tuân theo phân phối chuẩn và các phương sai không đồng nhất.

Sau khi biến đổi số liệu:

SAS code để biến đổi số liệu, phân tích và kiểm tra của VIDU19Transformation.SAS:

```
data WORK.VIDU19
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU19.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat CP best32. ;
informat MUA best32. ;
informat VSV comma32. ;
format CP best12. ;
format MUA best12. ;
format VSV comma12. ;
input
    CP
    MUA
    VSV
;
LOGVSV = LOG(VSV); ❶
PROC GLM;
CLASS CP MUA;
MODEL LOGVSV = CP MUA CP*MUA / SS4; ❷
OUTPUT OUT = DLMOI1 R = PHANDU1; ❸
RUN;

PROC UNIVARIATE NORMAL DATA = DLMOI1; ❹
VAR PHANDU1;
RUN;
```

Trong đó: ❶ Biến đổi cột số liệu VSV bằng làm logarit cơ số e, biến mới được tạo ra có tên LOGVSV; ❷ Mô hình phân tích số liệu tương tự như trên chỉ thay thế biến phụ thuộc VSV bằng LOGVSV; ❸ Tạo ra phần dư mới ❹ Để kiểm tra phân bố chuẩn của phần dư.

Kết quả từ SAS:

The GLM Procedure ❶

Class Level Information

Class	Levels	Values
CP	3	1 2 3
MUA	2	1 2

Number of observations	18
------------------------	----

The GLM Procedure

Dependent Variable: LOGVSV

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	33.69767569	6.73953514	2.92	0.0593
Error	12	27.65413730	2.30451144		
Corrected Total	17	61.35181299			

R-Square	Coeff Var		Root MSE	LOGVSV Mean	
0.549253	10.51349		1.518062	14.43918	
Source	DF	Type IV SS	Mean Square	F Value	Pr > F
CP	2	30.44788103	15.22394052	6.61	0.0116
MUA	1	0.45099922	0.45099922	0.20	0.6661
CP*MUA	2	2.79879545	1.39939772	0.61	0.5608

The UNIVARIATE Procedure

Variable: PHANDU1

Moments			
N	18	Sum Weights	18
Mean	0	Sum Observations	0
Std Deviation	1.27542697	Variance	1.62671396
Skewness	0.76337283	Kurtosis	-0.0547821
Uncorrected SS	27.6541373	Corrected SS	27.6541373
Coeff Variation	.	Std Error Mean	0.30062102

Basic Statistical Measures	
Location	Variability

Basic Statistical Measures			
Location		Variability	
Mean	0.00000	Std Deviation	1.27543
Median	-0.13569	Variance	1.62671
Mode	-0.92420	Range	4.53182
		Interquartile Range	1.30591

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > t	1.0000
Sign	M	-2	Pr >= M	0.4807
Signed Rank	S	-5.5	Pr >= S	0.8231

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.934769	Pr < W	❷ 0.2354
Kolmogorov-Smirnov	D	0.160141	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.066539	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.428513	Pr > A-Sq	>0.2500

Trong đó: ❶ Kết quả phân tích bằng thủ tục GLM sau khi số liệu được biến đổi bằng hàm logarit cơ số e; ❷ Xác suất của phần dư $P = 0,2354$ cho thấy sau khi biến đổi số liệu phần dư đã có phân bố chuẩn tức điều kiện của phép thử được thỏa mãn (số liệu có phân bố chuẩn và phương sai đồng nhất); ❸ Xác suất của yếu tố chế phẩm $P = 0,0116$ nên có thể kết luận rằng có sự sai khác về số lượng vi sinh vật giữa các chế phẩm khác nhau.

Lưu ý: trước khi biến đổi số liệu giả thiết H_0 được chấp nhận (chế phẩm không ảnh hưởng) và sau khi biến đổi số liệu giả thiết H_0 bị bác bỏ (chế phẩm ảnh hưởng đến số lượng vi sinh vật). Như vậy việc kiểm tra điều kiện của phép thử và tìm phương pháp biến đổi số liệu phù hợp có thể cho ta kết luận chính xác hơn.

2.6. PHÂN TÍCH PHƯƠNG SAI PHI THAM SỐ

Phân tích phương sai phi tham số được sử dụng khi số liệu không có phân bố chuẩn và cũng không thể biến đổi thành dạng phân phối chuẩn được. Muốn so sánh 2 công thức:

sử dụng phương pháp thống kê phi tham số Wilcoxon, nhiều công thức: Kruskal-Wallis và có thể sử dụng Friedman đối với thí nghiệm 2 nhân tố chéo nhau. Nếu trong phần thống kê tham số (đã nêu trên) tập trung vào so sánh các giá trị trung bình, thì trong phần thống kê phi tham số sẽ tiến hành so sánh các trung vị (median) với nhau. Ví dụ sau đây sẽ được sử dụng để minh họa sử dụng thống kê phi tham số trên phần mềm SAS.

Ví dụ 20: Thí nghiệm được tiến hành nhằm nghiên cứu ảnh hưởng của gossypol (một loại phenol tự nhiên có trong hạt bông) đến tăng khối lượng của vật nuôi. Có 5 mức gossypol khác nhau lần lượt là: 0; 0,04; 0,07; 0,1; 0,13. Số liệu thu được như sau:

0	0,04	0,07	0,10	0,13
228	186	179	130	154
229	229	193	87	130
218	220	183	135	130
216	208	180	116	118
224	228	143	118	118
208	198	204	165	104
235	222	114	151	112
229	273	188	59	134
233	216	178	126	98
219	198	134	64	100
224	213	208	78	104
220		196	94	
232			150	
200			160	
208			122	
232			110	
			178	

Giả thiết $H_0: \eta_1 = \eta_2 = \dots = \eta_t$ (các trung vị của quần thể bằng nhau)

Giả thiết $H_0: \eta_1 \neq \eta_2 \neq \dots \neq \eta_t$ (các trung vị của quần thể không bằng nhau)

Số liệu có thể nhập vào file VIDU20.XLS định dạng excel sau đó chuyển file VIDU20.TXT định dạng txt với 2 cột lần lượt là mức gossypol (MUC) và tăng khối lượng (P).

SAS CODE của VIDU20.SAS:

```
data WORK.VIDU20
;
%let _EFIERR_ = 0; /* set the ERROR detection macro variable */
infile 'D:\SAS2014\VIDU20.txt' delimiter='09'x MISSOVER DSD lrecl=32767
firstobs=2 ;
informat MUC best32. ;
```

```

informat P best32. ;
format MUC best12. ;
format P best12. ;

input
    MUC
    P
;

PROC NPAR1WAY ANOVA WILCOXON; ❶

CLASS MUC; ❷

VAR P; ❸

RUN;

```

Trong đó: ❶ Thủ tục thống kê phi tham số với lựa chọn ANOVA để cho kết quả phân tích phương sai và WILCOXON để kiểm định thống kê phi tham số; ❷ Yếu tố ảnh hưởng mức bổ sung gossypol (MUC) đến ❸ tăng khối lượng (P).

Kết quả từ SAS:

The NPAR1WAY Procedure ❶

Analysis of Variance for Variable P Classified by Variable MUC ❷		
MUC ❸	N ❹	Mean ❺
0	16	222.187500
0.04	11	217.363636
0.07	12	175.000000
0.1	17	120.176471
0.13	11	118.363636

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F ❻
Among	4	140082.986077	35020.74652	55.8143	<.0001
Within	62	38901.998997	627.45160		
Average scores were used for ties.					

Trong đó: ❶ Kết quả phân tích thống kê phi tham số; ❷ Kết quả phân tích phương sai; ❸ Các giá trị trung bình; ❹ Dung lượng mẫu tương ứng với ❺ các mức (công thức thí nghiệm) bổ sung gossypol; ❻ Xác suất ($P < 0,0001$) của yếu tố thí nghiệm trong kết quả phân tích phương sai.

The NPAR1WAY Procedure ❶

Wilcoxon Scores (Rank Sums) for Variable P ② Classified by Variable MUC					
MUC ⑤	N ④	Sum of Scores	Expected Under H0	Std Dev Under H0	③ Mean Score
0	16	890.50	544.0	67.978966	55.656250
0.04	11	555.00	374.0	59.063588	50.454545
0.07	12	395.50	408.0	61.136622	32.958333
0.1	17	275.50	578.0	69.380741	16.205882
0.13	11	161.50	374.0	59.063588	14.681818
Average scores were used for ties.					

Kruskal-Wallis Test ⑥	
Chi-Square	52.6656
DF	4
Pr > Chi-Square	<.0001 ⑦

Trong đó: **①** Kết quả phân tích thống kê phi tham số; **②** Kết quả phân tích xếp hạng từ WILCOXON option của câu lệnh nêu trên; **③** Trung bình của thứ hạng; **④** Dung lượng mẫu tương ứng **⑤** Các mức (công thức thí nghiệm) bổ sung gossypol; **⑥** Kết quả kiểm định Kruskal-Wallis; Với **⑦** xác suất $P < 0,0001$ nên có thể kết luận: Các trung vị không bằng nhau hay nói cách khác tăng khối lượng trung bình có sự sai khác giữa các mức bổ sung.

CÂU HỎI ÔN TẬP

Một thí nghiệm được tiến hành trên 40 con lợn Piétrain nhằm nghiên cứu khả năng sinh trưởng của giống lợn này. Tại thời điểm bắt đầu thí nghiệm từng cá thể được đánh số tai (ST), xác định giới tính (GT, đực hoặc cái), kiểu gen halothane (GEN, CC hoặc CT), cân khối lượng (P0, kg) và số ngày tuổi tương ứng của từng cá thể (TUOI0, ngày). Tại thời điểm kết thúc thí nghiệm cân khối lượng của từng cá thể (P1, kg), siêu âm để xác định độ dày mỡ lưng (ML, mm), độ dày cơ thăn (CT, mm), tỷ lệ nạc (NAC, %) và số ngày tuổi tương ứng của từng cá thể (TUOI1, ngày). Số liệu thu được trình bày ở bảng dưới:

ST	GT	GEN	TUOI0	P0	TUOI1	P1	ML	CT	NAC	ST	GT	GEN	TUOI0	P0	TUOI1	P1	ML	CT	NAC
1	cai	CC	55	13	195	71.5	9.1	48.4	61.35	21	duc	CC	55	15	195	91.5	7.5	52	63.87
2	cai	CC	55	13.4	195	81	8.8	50.3	62.10	22	duc	CC	59	17	198	92.5	7.1	45.5	62.81
3	cai	CC	58	13.2	198	80	6.5	51	64.70	23	duc	CC	61	16.6	200	65	5.2	43.6	64.39
4	cai	CC	58	13.8	198	81	8.1	49.1	62.57	24	duc	CC	61	14	200	72.5	8.4	40.7	60.33
5	cai	CC	59	17.4	199	73	8.4	44.6	61.22	25	duc	CC	61	14	200	73.5	7.5	45.1	62.29
6	cai	CC	61	11	200	70.5	7.1	46.5	63.03	26	duc	CC	61	14	201	80	7.1	50.1	63.86
7	cai	CC	61	16.6	200	76.5	6.5	47	63.79	27	duc	CC	61	17	200	85.5	8.4	48.8	62.18
8	cai	CC	61	19.6	200	77.5	6.5	51.7	64.86	28	duc	CC	61	13.4	200	92	8.8	52.3	62.56
9	cai	CC	61	14.6	200	80	8.4	47.8	61.95	29	duc	CC	61	14.4	200	95.5	9.8	48	60.51
10	cai	CC	70	18.6	197	77	8.8	50.3	62.10	30	duc	CC	70	23.8	197	90.5	8.1	48.8	62.50
11	cai	CC	73	18.6	200	66.5	6.8	43.6	62.69	31	duc	CC	73	21.6	200	70.5	7.1	48.8	63.56
12	cai	CC	73	22.1	200	81	8.4	52.4	63.01	32	duc	CC	73	26.6	200	91.5	9.1	56.5	63.21
13	cai	CT	58	14.8	197	67.5	7.1	49.4	63.70	33	duc	CC	73	23.6	201	97	10.7	54	60.94
14	cai	CT	59	15.4	198	72.5	6.5	48.4	64.11	34	duc	CT	57	18.4	184	75	6.5	42.9	62.85
15	cai	CT	59	13.2	199	75.5	8.1	57.2	64.43	35	duc	CT	57	18	185	79.5	7.8	49.1	62.89
16	cai	CT	59	18.4	198	84.5	5.8	47.2	64.57	36	duc	CT	58	12.8	197	71	5.2	41.6	63.93
17	cai	CT	59	16.2	198	88	7.8	54.3	64.08	37	duc	CT	58	15.2	198	76	8.8	41.9	60.18
18	cai	CT	71	20.6	199	69.5	8.5	40.3	60.13	38	duc	CT	59	14.6	199	72	6.8	41.3	62.16
19	cai	CT	71	22.6	198	71.5	6.5	43.5	62.98	39	duc	CT	59	17.6	199	96	8.4	48.8	62.18
20	duc	CC	55	11.2	194	65	7.1	46.5	63.03	40	duc	CT	61	13	200	78.5	7.1	47.2	63.20

Yêu cầu:

Tạo thêm biến mới “tăng khối lượng trung bình (g/ngày) - ADG” vào bộ số liệu nêu trên theo công thức: $ADG = 1000 \cdot (P1 - P0) / (TUOI1 - TUOI0)$.

1. Tính các tham số thống kê mô tả của các chỉ tiêu nghiên cứu nêu trên (tính chung; tính theo giới tính, kiểu gen; tính theo giới tính - kiểu gen).
2. Đánh giá ảnh hưởng của kiểu gen và giới tính bằng phép phân tích phương sai (ANOVA). Có 2 mô hình xử lý gồm: (1) phân tích từng yếu tố (kiểu gen hoặc giới tính) riêng biệt như mô hình thiết kế thí nghiệm một yếu tố hoàn toàn ngẫu nhiên và (2) phân tích 2 yếu tố (kiểu gen và giới tính) đồng thời như mô hình thiết kế thí nghiệm có 2 yếu tố.
3. Phân tích hiệp phương sai (ANCOVA) với hiệp biến là khối lượng bắt đầu thí nghiệm (P0) và 2 yếu tố cố định (kiểu gen hoặc giới tính). Các yếu tố này ảnh hưởng như thế nào đến tăng khối lượng trung bình/ngày. So sánh với kết quả ở câu 2.
4. Tính hệ số tương quan giữa các tính trạng nghiên cứu.
5. Xây dựng mô hình hồi quy tuyến tính để xác định tỷ lệ nạc thông qua độ dày mỡ lưng và cơ thăn.

Chương 3

CÔNG BỐ KẾT QUẢ NGHIÊN CỨU

Một công trình nghiên cứu khoa học chưa được coi là hoàn thành khi kết quả chưa được công bố. Các nhà khoa học công bố kết quả nghiên cứu để chia sẻ những thông tin, kiến thức và hiểu biết mới trong một lĩnh vực nhất định. Nội dung chính của chương này là hướng dẫn nghiên cứu sinh trong quá trình “từ nghiên cứu đến công bố” gồm: (1) Cách trình bày kết quả nghiên cứu, (2) Cách chuẩn bị và thuyết trình báo cáo khoa học, (3) Cách viết bài báo khoa học và (4) Cách viết Luận án tiến sĩ. Một số vấn đề về văn phong khoa học và bản quyền trong công bố khoa học cũng được đề cập tới trong chương này.

3.1. TRÌNH BÀY KẾT QUẢ NGHIÊN CỨU

Đối với một công bố khoa học gốc, nhà khoa học công bố kết quả nghiên cứu của chính mình. Kết quả nghiên cứu có thể công bố dưới nhiều hình thức, phổ biến nhất vẫn là qua các báo cáo trình bày tại các hội nghị khoa học hay các bài báo khoa học xuất bản trên các tạp chí. Một công bố khoa học như vậy có thuyết phục hay không chính là ở phần trình bày số liệu vì đó là phần cung cấp bằng chứng thực tế khách quan để trả lời câu hỏi hay giả thuyết nghiên cứu. Do vậy, sau khi có kết quả phân tích thống kê, công việc tiếp theo của người nghiên cứu là tóm tắt và lựa chọn cách trình bày các kết quả đó trong công bố khoa học theo các hình thức và trình tự hợp lý sao cho người đọc/người nghe dễ hiểu. Đó là một thách thức lớn đối với nghiên cứu sinh.

Tùy theo loại kết quả phân tích số liệu mà người nghiên cứu có thể trình bày kết quả theo 1 trong 3 dạng sau: bảng số liệu (table), biểu đồ (graphics) hay văn viết (text). Bảng số liệu thường sử dụng cho những trường hợp cần tóm lược những thông tin mang tính chính xác cao. Biểu đồ có thể sử dụng để chuyển tải những thông điệp về mức độ ảnh hưởng và xu hướng biến thiên của dữ liệu, nhất là khi dung lượng dữ liệu lớn. Văn viết chỉ sử dụng cho những dữ liệu rất đơn giản, khó thể hiện được tất cả những xu hướng và dao động của dữ liệu.

3.1.1. Bảng số liệu

a. Yêu cầu chung cho các bảng số liệu

Bảng số liệu là hình thức phổ biến nhất để trình bày kết quả nghiên cứu. Bảng số liệu phải thỏa mãn được một số yêu cầu sau:

- Bảng số liệu phải tự giải thích được: Khi nhìn vào bảng số liệu người đọc phải hiểu được những thông tin trong đó mà không cần giải thích gì thêm. Do vậy, bảng số liệu thường phải chứa các thành phần cấu trúc sau đây:

- + Số và tên bảng (title);

- + Tiêu đề cột (column label);
- + Tiêu đề hàng (row label);
- + Phần thân chính chứa số liệu (data);
- + Chú thích cuối bảng;
- + Ranh giới giữa các phần.

Mỗi bảng số liệu phải có tên, có đánh số và được đặt phía trên của bảng. Tên bảng phải phản ánh được những thông tin cơ bản để người đọc biết bảng số liệu nói lên điều gì. Cột và dòng phải có tiêu đề và được sắp xếp một cách logic, có thứ tự. Thông thường các cột chứa tên các biến độc lập/lô thí nghiệm, còn các hàng chứa tên các biến phụ thuộc (các chỉ tiêu theo dõi thí nghiệm như khối lượng, tăng khối lượng, tiêu tốn thức ăn...). Các chỉ tiêu phải có đơn vị đo lường (thường ghi trong ngoặc đơn hoặc sau dấu phẩy sau tên của chỉ tiêu). Ví dụ: Khối lượng đầu kỳ (kg). Tất cả những viết tắt và ký hiệu thống kê trong bảng phải được giải thích ngắn gọn ở phần chú thích phía dưới mỗi bảng số liệu. Nếu số liệu là các giá trị thống kê như “ 65 ± 12 ” thì tác giả phải định nghĩa các con số là Mean $\pm$ SD, Mean $\pm$ SE hay Median $\pm$ SE?

- Số liệu phải đơn giản. Không nên để quá nhiều số chữ số thập phân trong mỗi số liệu. Trong hầu hết các trường hợp, số chữ số thập phân của số liệu trong bảng tương đương với số chữ số thập phân của đơn vị đo lường gốc.

- Sử dụng đúng các tham số thống kê. Số trung bình thường là những giá trị chính trong mỗi bảng số liệu. Kèm theo phải có các ký tự (như ^a, ^b, ^c) để làm chỉ số đánh dấu kết quả so sánh thống kê. Ngoài ra còn có các tham số thống kê khác như độ lệch chuẩn (SD), sai số chuẩn (SE/SEM) và trị số P. Khi dùng thống kê mô tả để phản ánh sự biến thiên của quần thể thì dùng SD, còn khi dùng thống kê so sánh để tìm sự sai khác giữa các nghiệm thức thì dùng SE hay SEM. Về trị số P, khi $P > 0,05$ thì được coi là giả thiết vô hiệu (H_0) là đúng và ngược lại khi $P < 0,05$ thì chấp nhận đối thiết (H_1), tức là chấp nhận sự sai khác có ý nghĩa. Trước đây do chưa có máy tính nên P thường chỉ được dùng ở 3 mức 0,05, 0,01 và 0,001. Tuy nhiên, ngày nay các tạp chí khoa học thường yêu cầu tác giả trình bày giá trị chính xác của trị số P. Mặc dù vậy, khi kết quả phân tích thống kê cho $P = 0,0000$ thì lại nên ghi là $P < 0,001$.

- Những số liệu quan trọng cần được nhấn mạnh hay làm nổi bật. Nhấn mạnh bằng cách tô đậm hay gạch chân (nhất là khi trình bày PowerPoint) những số liệu quan trọng như những giá trị trung bình sai khác nhau có ý nghĩa.

- Chỉ dùng bảng khi số liệu có thể phản ánh được kết quả nghiên cứu hiệu quả hơn biểu đồ hay văn viết. Bảng không được sử dụng khi có ít số liệu (<6), thay vào đó nên trình bày ở dạng văn viết; và cũng không trình bày bảng khi có quá nhiều số liệu (>40), thay vào đó nên dùng biểu đồ. Nên sử dụng bảng khi số liệu có các đặc trưng sau:

(1) Thể hiện tính hệ thống; (2) Rõ ràng, chính xác; và (3) Người đọc dễ hiểu, dễ thấy được sự khác nhau và rút ra nhiều kết luận về mối quan hệ giữa các số liệu với nhau.

- Dạng thức của bảng số liệu phải phù hợp với hình thức công bố kết quả nghiên cứu. Sự khác nhau được thể hiện khi trình bày bảng số liệu trong bài báo khoa học và trong báo cáo khoa học bằng PowerPoint. Trong một bài báo khoa học bảng số liệu có thể phức tạp, nhưng bảng số liệu trong PowerPoint nên đơn giản, đi thẳng vào vấn đề. Một sai lầm hay gặp phải của nhiều người là sao chép bảng số liệu từ bài báo hay văn bản báo cáo (dạng word) sang bản chiếu PowerPoint. Nguyên tắc số 1 của trình bày số liệu trên PowerPoint là đơn giản và dễ hiểu. Do đó, một bảng số liệu trong PowerPoint không nên có hơn 4 cột và 5 dòng.

Mỗi tạp chí chuyên ngành có thể có yêu cầu đặc thù riêng về hình thức trình bày bảng số liệu. Do đó, tác giả phải nắm được quy định của Tạp chí định gửi đăng bài báo mà định dạng các bảng số liệu cho phù hợp.

b. Các loại bảng số liệu

Các bảng số liệu có thể phân thành 4 nhóm chính là: bảng danh sách, bảng mô tả, bảng so sánh và bảng phân tích đa biến.

- Bảng danh sách

Đây là dạng bảng đơn giản nhất, thường chỉ có một vài cột và nhiều dòng để trình bày một danh sách dữ liệu. Ví dụ:

Bảng 3.1. Các yếu tố nội tại liên quan đến tính khả thi của chăn nuôi thỏ nông hộ

Yếu tố	Mô tả
Thức ăn	1) Các nguồn lực tại chỗ (thức ăn xanh, diện tích đất)? 2) Nhu cầu mua thức ăn thương phẩm?
Con giống	1) Có sẵn giống địa phương không? 2) Phương pháp nhân giống phù hợp nhất?
Thú y	1) Các bệnh thường gặp và cách điều trị?
Chuồng trại	1) Nguyên liệu rẻ tiền, sẵn có? 2) Nông dân có có khả năng mua nguyên liệu làm chuồng không?

- Bảng mô tả đơn giản

Bảng dạng này thường dùng để mô tả kết quả nghiên cứu có một nhóm đối tượng duy nhất. Số liệu có thể thuộc dạng biến liên tục hay biến rời rạc. Biến liên tục thường được mô tả bằng các tham số thống kê mô tả như giá trị trung bình, độ lệch chuẩn, sai số chuẩn. Biến rời rạc (hay biến phân loại) thường được mô tả bằng con số nguyên (số lượng) và số phần trăm (tỷ lệ). Ví dụ:

Bảng 3.2. Số lợn sơ sinh qua 7 lứa tại Trung tâm Giống lợn chất lượng cao, Học viện Nông nghiệp Việt Nam (n = 1119)

Lứa	Số con đẻ ra (con)	Tần suất (%)
1	337	30,12
2	275	24,58
3	213	19,03
4	137	12,24
5	86	7,69
6	49	4,38
7	22	1,97

- Bảng so sánh

Bảng số liệu có mục đích so sánh là dạng bảng rất phổ biến trong các công bố khoa học. Bảng thường có nhiều hơn 2 cột. Các nhóm so sánh nên thể hiện ở các cột, còn các chỉ tiêu so sánh nên để ở các hàng. Mỗi ô số liệu ngoài việc chứa các giá trị chính của mỗi nhóm còn có thể có thêm kết quả phân tích thống kê như khoảng tin cậy, ký tự đánh dấu mức độ sai khác... Ví dụ:

Bảng 3.3. Tỷ lệ thịt và xương trong thân thịt của dê theo phẩm giống và chế độ nuôi

	Phẩm giống		Chế độ nuôi dưỡng		SEM
	Dê lai F1 (n = 6)	Dê địa phương (n = 6)	Cải tiến (n = 6)	Truyền thống (n = 6)	
Khối lượng thân thịt (kg)	8,38 ^a	6,25 ^b	7,83	6,80	0,35
Tỷ lệ thịt (%)	69,79	70,28	70,72	69,35	0,87
Tỷ lệ xương (%)	30,21	29,72	29,28	30,65	0,87

Ghi chú: Trong cùng một hàng, những giá trị trung bình của hai phẩm giống hay hai chế độ nuôi có chữ cái khác nhau thì sai khác ở mức ý nghĩa $P < 0,05$; SEM: Sai số của số trung bình.

- Bảng phân tích đa biến

Trong nhiều nghiên cứu chăn nuôi một số chỉ tiêu nghiên cứu được phân tích thống kê theo mô hình đa biến (multivariable analysis). Trong những phân tích này với mỗi biến phụ thuộc (chỉ tiêu nghiên cứu) có nhiều biến độc lập (yếu tố tác động). Trong trường hợp này kết quả phân tích thường được thể hiện dưới dạng phương trình hồi quy đa biến và/hoặc bảng phân tích đa biến. Ví dụ, bảng 3.4 thể hiện cách trình bày kết quả phân tích mô hình phân tích hồi quy đa biến: $y_i = a + b_1 \cdot x_{1i} + b_2 \cdot x_{2i} + \dots + b_k \cdot x_{ki} + e_i$

Bảng 3.4. Kết quả phân tích hồi quy đa biến...

Mô hình	Hệ số hồi quy (b)	SE	t	P
Điểm chắn (Intercept)				
Biến 1				
Biến 2				
...				
Biến n				
Số lượng quan sát				
R^2				
R^2 hiệu chỉnh				
F				

3.1.2. Biểu đồ

a. Nguyên tắc soạn biểu đồ

Khi trình bày dữ liệu bằng biểu đồ cần phải tuân thủ các nguyên tắc sau:

- Biểu đồ phải tự giải thích được. Cũng như đối với bảng số liệu, yêu cầu với một biểu đồ là khi nhìn vào người xem tự hiểu được thông tin, không cần giải thích gì thêm. Muốn vậy, trước tiên mỗi biểu đồ phải có một cái tên phản ánh được thông điệp chính mà tác giả muốn chuyển tải. Tên biểu đồ thường được đặt ở phía dưới biểu đồ. Ngoài ra, các trục của biểu đồ cũng phải có tiêu đề, đơn vị tính với cự ly thích hợp. Đồng thời trên biểu đồ phải có các chú giải cần thiết khác.

- Nói lên bản chất biến động của dữ liệu. Biểu đồ không chỉ biểu diễn các giá trị trung bình đơn thuần mà cần thể hiện được cả những dao động của dữ liệu trong mỗi nhóm hay giữa các nhóm đối tượng cần so sánh. Chẳng hạn như một biểu đồ thanh (bar chart) trình bày số trung bình về khoảng cách lựa đề giữa nhóm can thiệp và nhóm đối chứng là chưa phản ánh được sự biến động của dữ liệu, bởi vì khoảng cách lựa đề của các cá thể trong mỗi nhóm dao động khá lớn. Do đó, một biểu đồ có ý nghĩa hơn là biểu đồ hộp (whisker plot/box chart), hay tốt hơn nữa là biểu đồ hộp cộng với những số liệu cho từng cá thể trong mỗi nhóm.

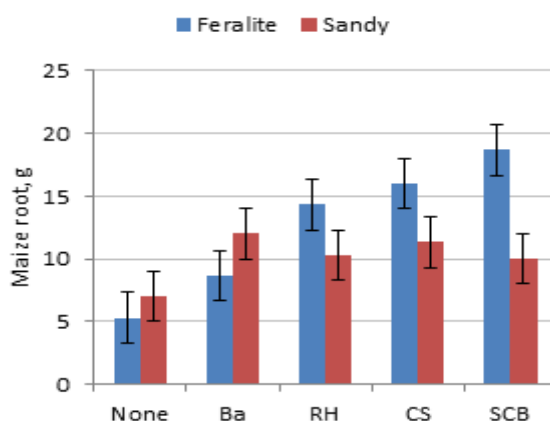
- Trình bày dữ liệu càng nhiều càng tốt so với lượng mực in cần dùng. Những biểu đồ bánh (pie chart) thường cho thấy số liệu rất ít so với lượng mực in. Vì thế, loại biểu đồ này ít được sử dụng trong các công bố khoa học.

b. Các loại biểu đồ

Có nhiều dạng biểu đồ và mỗi dạng chỉ có thể áp dụng cho một tình huống phù hợp. Các dạng hình thường được sử dụng trong công bố kết quả thí nghiệm chăn nuôi gồm: biểu đồ cột (colume chart), biểu đồ thanh (bar chart), biểu đồ phân bố tần suất (frequency histogram), biểu đồ hộp (box chart), biểu đồ phân tán (scatter plot), biểu đồ đường (line chart), biểu đồ hình bánh (pie chart), ...

- Biểu đồ cột và thanh

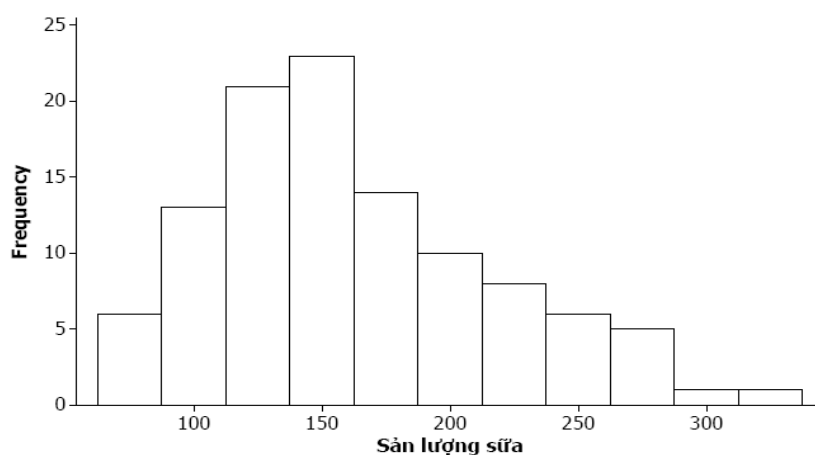
Biểu đồ cột (colume chart) và biểu đồ thanh (bar chart) được sử dụng để so sánh sự khác biệt về quy mô khối lượng của một số đối tượng nào đó; thể hiện tương quan về độ lớn về các đại lượng. Các cột đơn thể hiện các đại lượng khác nhau (có thể đặt cạnh nhau), ta có biểu đồ cột - gộp nhóm. Nếu biến phân tích là biến liên tục có thể thêm sai số chuẩn (SE) vào cuối cột hay thanh để biểu thị độ dao động của giá trị trung bình (Biểu đồ 3.1).



Biểu đồ 3.1. Ảnh hưởng của loại đất (Feralite và Sandy) và nguồn biochar (RH: trâu; Ba: Tre; CS: vỏ dừa; SCB: bã mía) đến sinh khối rễ cây ngô sau 35 ngày sinh trưởng

- Biểu đồ phân bố tần suất (frequency histogram)

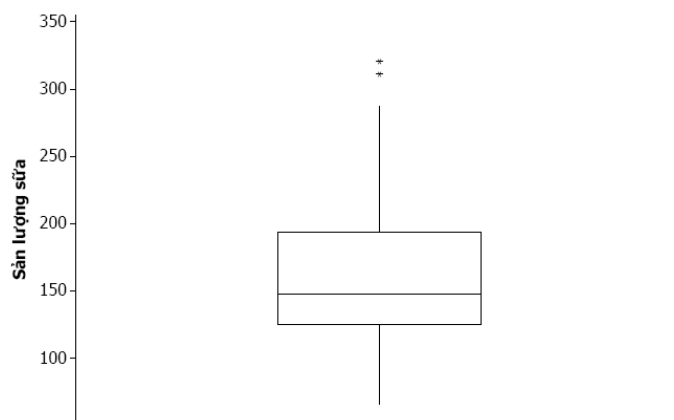
Biểu đồ phân bố tần suất (còn gọi là tổ chức đồ) thường được dùng để biểu thị sự phân bố của một biến liên tục thể hiện qua số liệu đo của các cá thể phân bố dọc theo trục của biến (Biểu đồ 3.2). Tần suất (trục y) có thể là trị số tuyệt đối (số đếm) hoặc tương đối (phần trăm hoặc tỷ lệ của mẫu). Trình bày bằng đồ thị tần suất cần thiết khi mô tả quần thể.



Biểu đồ 3.2. Phân bố tần suất sản lượng sữa của dê Bách Thảo (kg/chu kỳ)

- Biểu đồ hộp (box chart/whisker plot)

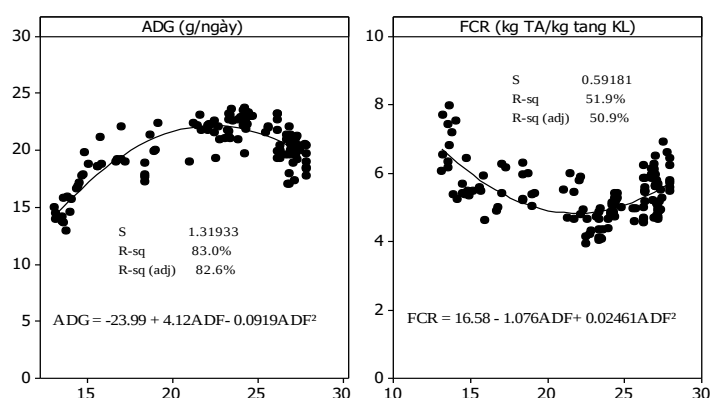
Biểu đồ hộp là một dạng biểu đồ khác để thể hiện sự phân bố của dữ liệu, đồng thời cung cấp những tham số thống kê mô tả như số trung bình, trung vị, tối đa, tối thiểu mà biểu đồ phân bố có thể không thể hiện được (Biểu đồ 3.3).



Biểu đồ 3.3. Phân bố tần suất sản lượng sữa dê Bách Thảo (kg/chu kỳ)

- Biểu đồ phân tán (scatter plot)

Biểu đồ phân tán (hay tán xạ) được sử dụng để trình bày sự phân bố và mối quan hệ giữa các số liệu của hai biến. Trong đó, các giá trị là các chấm phân bố và mối quan hệ được thể hiện bằng đường hồi quy (Biểu đồ 3.4). Nếu có mối quan hệ giữa các biến, biến độc lập nên chọn là trục x và biến phụ thuộc là trục y. **ví dụ**, chiều cao vây phụ thuộc vào độ tuổi, như vậy chiều cao vây là biến phụ thuộc được biểu diễn trên trục y và tuổi là biến độc lập là trục x. Đôi khi có trường hợp khó xác định được biến nào là biến phụ thuộc hay biến độc lập. Trong trường hợp này, không xác định được ảnh hưởng của biến nào đối với biến nào thì trình bày trong mối quan hệ tự chọn.

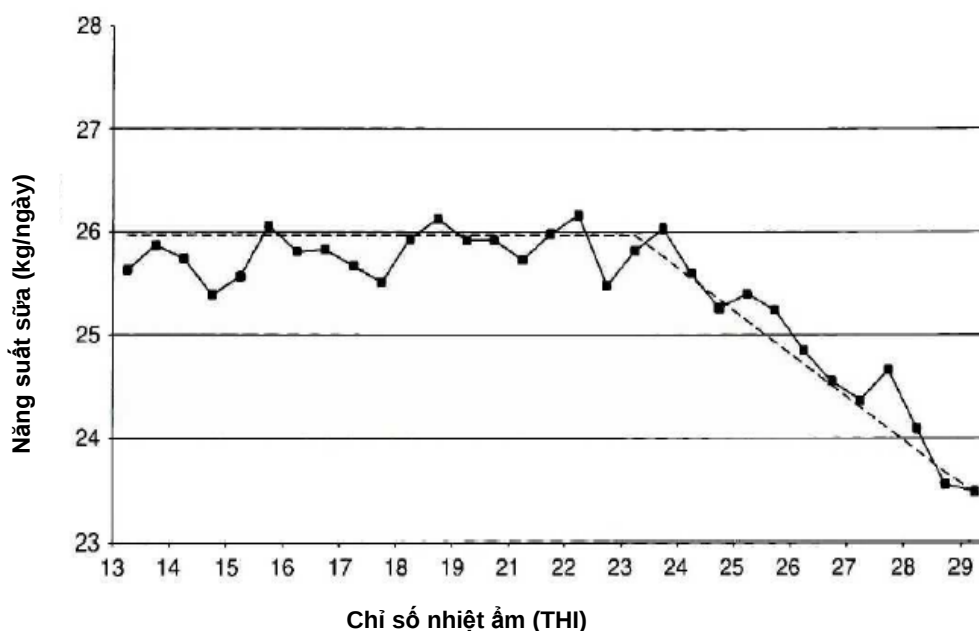


Biểu đồ 3.4. Hồi quy giữa tốc độ tăng khối lượng (ADG) và hệ số chuyển hoá thức ăn (FCR) của thỏ New Zealand với hàm lượng ADF trong khẩu phần

Ghi chú: Hàm lượng ADF trong khẩu phần (% vật chất khô)

- Biểu đồ đường (line chart)

Biểu đồ đường được trình bày khi các giá trị của biến phụ thuộc là chuỗi liên tục như thể trọng, năng suất sữa... Các giá trị là các điểm được nối với nhau bởi đường thẳng hoặc đường cong diễn tả xu hướng biến động của biến phụ thuộc theo biến độc lập (Biểu đồ 3.5). Có thể trình bày đường biểu diễn của nhiều biến phụ thuộc (vào cùng 1 biến độc lập) trên cùng một biểu đồ.

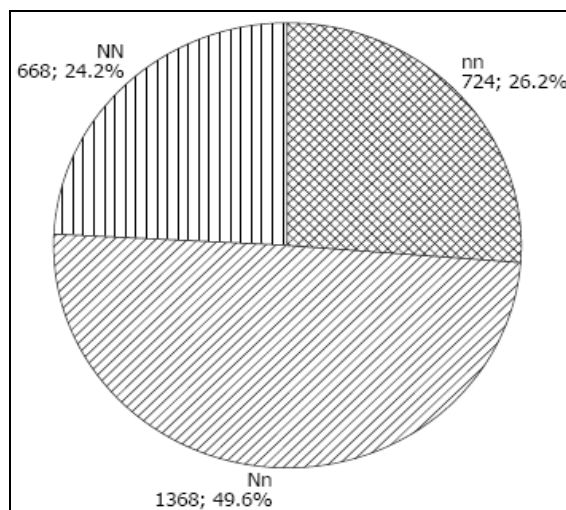


Biểu đồ 3.5. Ảnh hưởng của chỉ số nhiệt ẩm (THI) lên năng suất sữa của bò

- Biểu đồ hình bánh (pie chart)

Được sử dụng để trình bày mối quan hệ tỷ lệ so sánh theo phần trăm tổng của các số liệu khác nhau theo một biến phân loại (Biểu đồ 3.6). Loại biểu đồ này ít được dùng trong nghiên cứu khoa học. Khi trình bày các số liệu bằng biểu đồ hình bánh nên tuân theo các qui tắc sau:

- + Tổng số các số liệu có giá trị tổng không đổi (thường 100%)
- + Các giá trị có sự khác biệt tương đối lớn (có ý nghĩa). Khi các giá trị bằng nhau thì không nên trình bày bằng biểu đồ này
- + Mỗi phần chia của hình (mỗi phần tương ứng với một giá trị) nên được chú thích
- + Số phần chia tương đối nhỏ (thông thường là từ 3-7 phần) và không vượt quá 7



Biểu đồ 3.6. Tần số kiểu gen Halothane của lợn sơ sinh Pietrain (n = 2760)

3.1.3. Văn viết mô tả số liệu

Không phải tất cả các số liệu phân tích hay kết quả nghiên cứu đều phải trình bày ở dạng bảng hay biểu đồ. Những số liệu đơn giản, chỉ nên trình bày ở dạng văn viết và cho các số liệu vào trong ngoặc. Ví dụ, “Số con đẻ ra trên lứa của lợn Móng Cái ($11,3 \pm 0,8$) cao hơn lợn Landrace ($10,1 \pm 0,7$), ($P < 0,01$)”. Hoặc với một kết quả khảo sát chỉ có một vài con số kiểu như trong một đàn vật nuôi có 45% đực và 55% cái thì không cần thiết phải dùng bảng hay biểu đồ vì có thể viết ngắn gọn hơn. Ví dụ, có thể viết “Theo kết quả khảo sát thì đàn vật nuôi có 45% đực” (đương nhiên là có 55% cái, cũng không cần phải viết).

Mặt khác, vì dữ liệu là “bột” để “gột” thành một công bố khoa học nên đối với các số liệu quan trọng thì ngoài việc trình bày trong các bảng hay biểu đồ (gọi chung là *bảng biểu*) chúng cũng cần được diễn giải bằng lời để hướng dẫn cho người đọc dễ theo dõi. Tác giả phải viết sao cho người đọc không cần nhìn vào bảng biểu vẫn có thể nắm được nội dung và ý nghĩa của kết quả nghiên cứu. Do vậy, khi mô tả một bảng biểu, tác giả cần tìm những câu từ rõ ràng và súc tích nhất mà *không cần viết lại những con số đã có trong bảng biểu* để trả lời những câu hỏi quan trọng sau:

1. Bảng biểu đó nói về nội dung gì?
2. Những thông tin chính trên bảng biểu đó là gì?
3. Thông điệp chính của bảng biểu đó là gì?

Ví dụ, để mô tả biểu đồ 3.5 ở trên có thể viết theo hai cách như sau:

Cách 1: “*Biểu đồ 3.5 cho thấy tần số xuất hiện của 3 kiểu gen Halothane (NN, Nn và nn) ở lợn sơ sinh Pietrain. Trong tổng số 2.760 lợn con theo dõi có 668 con (chiếm 24,2%) mang kiểu gen NN, 1.368 con (chiếm 49,6%) mang kiểu gen Nn và 724 con (chiếm 26,2%) mang kiểu gen nn. Như vậy, lợn con mang kiểu gen Nn chiếm tỷ lệ cao*

nhất.” Cách viết này rất phổ biến hiện nay trong các luận án, nhưng thực ra đó chỉ là cách mất thời gian “phiên dịch” lại biểu đồ sang văn viết mà không nói thêm được thông tin gì, không nêu được đâu là thông điệp chính, nên không cần thiết vì người đọc đã có thể dễ dàng đọc và thấy được những gì vừa viết lại ngay trên biểu đồ rồi.

Cách 2: “Theo biểu đồ 3.5, tần số xuất hiện của 3 kiểu gen *Halothane* NN, Nn và nn ở lợn sơ sinh Pietrain xấp xỉ với tỷ lệ 1:2:1”. Cách viết này ngắn gọn, khách quan, không lặp lại các con số trong biểu đồ mà vẫn đáp ứng được 3 câu hỏi nêu trên, cụ thể là biểu đồ 3.5 có (1) nội dung về “...kiểu gen *Halothane* ... ở lợn sơ sinh Pietrain...” với (2) thông tin chính là “..., tần số xuất hiện của 3 kiểu gen... NN, Nn và nn...” và (3) thông điệp chính là “..., tần số xuất hiện... xấp xỉ với tỷ lệ 1:2:1”. Đây là một thông điệp để làm tiền đề cho phần *Thảo luận* có thể đặt câu hỏi chẳng hạn như “*Phải chăng đàn lợn bố mẹ Pietrain ở đây có cùng kiểu gen Nn nên mới có tỷ lệ phân ly như vậy ở đời con?* Nếu tác giả chứng minh được không phải vậy (đã biết chính xác kiểu gen của đàn bố mẹ chẳng hạn) nhưng cũng chưa tìm được lý do nào khác để trả lời xác đáng thì đó có thể lại là cơ sở để hình thành nên một câu hỏi nghiên cứu mới.

3.2. BÁO CÁO KHOA HỌC

Báo cáo khoa học là bài báo cáo trong hội nghị khoa học về kết quả thực hiện một công trình nghiên cứu. Đó là một bài thuyết trình trước một nhóm người trong một khoảng thời gian cho phép từ 15 đến 30 phút. Khoảng thời gian đó không đủ để trình bày đầy đủ mọi thứ như trong bài báo viết. Do đó, để một báo cáo khoa học có chất lượng tốt thì ngoài nội dung khoa học tốt cần được thiết kế thông điệp một cách hoàn chỉnh thông qua các bản chiếu (slides) cùng với một phong cách trình bày làm sao để làm nổi bật được các nội dung chính yếu và quan trọng nhất.

Có nhiều kiểu thuyết trình báo cáo khoa học khác nhau. Tuy nhiên, ở đây chỉ đề cập đến phương pháp chuẩn bị và cách thuyết trình một báo cáo khoa học với sự hỗ trợ của phần mềm PowerPoint vì việc sử dụng phần mềm PowerPoint khi thuyết trình báo cáo khoa học ngày càng trở nên phổ biến trong thời đại công nghệ thông tin hiện nay. Đây cũng là phương tiện mà mỗi nghiên cứu sinh (NCS) phải dùng đến trong các buổi hội thảo chuyên môn và đặc biệt là trong buổi “Bảo vệ Luận án tiến sĩ”. Hơn nữa, người báo cáo cũng cần biết được cách thức trả lời chất vấn của khán giả sau khi trình bày báo cáo.

3.2.1. Chuẩn bị báo cáo

a. Tìm hiểu

Trước khi chuẩn bị bài báo cáo PowerPoint cần phải tìm hiểu thông tin cơ bản về buổi báo cáo để cấu trúc bài báo cáo cho phù hợp với tình hình thực tế. Những thông tin cần có là: thời lượng được phép của báo cáo, thời gian biểu trình bày trong chương trình, không gian hội trường, thành phần tham dự, ai là chủ tọa, ai báo cáo trước mình...

Biết được thời lượng để quyết định số bản chiếu cần thiết vì mỗi bản chiếu thường được trình bày trung bình trong vòng 1 phút. Biết được giờ cụ thể để chuẩn bị cách nói. Nếu là báo cáo vào đầu buổi sáng, khi khán giả vẫn còn hào hứng, thì có cách nói bình thường. Nếu nói vào đầu giờ buổi chiều, khán giả có thể rất dễ buồn ngủ và không tập trung thì phải nói sao cho họ không ... ngủ. Biết được kích thước hội trường để pha màu bản chiếu cho phù hợp. Nếu hội trường rộng nên để màu nền đậm và chữ màu sáng; ngược lại, nếu hội trường nhỏ hẹp thì dùng màu nền sáng và chữ màu đậm. Biết được thành phần khán giả để điều chỉnh cách nói. Nếu người nghe chủ yếu là những nhà khoa học trình độ cao thì nên chọn cách nói hàn lâm. Nếu người nghe là đồng môn thì nên nói một cách thân mật. Nguyên tắc là không được để người nghe cảm thấy bị xúc phạm. Tìm hiểu ai là chủ tọa, chức danh khoa học của (những) người đó là gì, để tiện việc xưng hô và nói xã giao. Biết được ai báo cáo trước mình để nói câu xã giao phù hợp trước khi bắt đầu trình bày báo cáo của mình (nếu cần).

b. Thiết kế bài báo cáo

** Bố cục bài báo cáo*

Nguyên tắc bố cục chung của một bài báo cáo PowerPoint là: “Tell what will be talked, talk what to be talked, tell what have been talked” (nói với khán giả những gì sẽ được trình bày, trình bày những gì cần nói, tóm tắt lại những gì đã trình bày). Vì vậy, sau bản chiếu tựa đề đầu tiên cần có 1 bản chiếu giới thiệu NỘI DUNG bài báo cáo (what will be talked), tiếp theo là các bản chiếu trình bày nội dung chính của báo cáo (what to be talked), và cuối cùng có 1 bản chiếu tóm tắt lại những nội dung chính đã được trình bày (what have been talked) và/hoặc những kết luận rút ra từ những nội dung đã thể hiện trong các bản chiếu trước đó.

** Tên bài báo cáo*

Tên bài của một báo cáo khoa học cũng giống như là một dòng chữ quảng cáo. Một cái tên hấp dẫn có khả năng thu hút sự chú ý của khán giả trong hội nghị ngay từ đầu. Để đạt được mục đích đó, tên báo cáo phải có đầy đủ thông tin cơ bản về bài báo cáo, súc tích, nhưng không quá phức tạp mà cũng không quá chung chung. Tên phức tạp sẽ làm cho người nghe không thấy hấp dẫn. Tên chung chung làm người nghe không có động cơ để theo dõi nên không tập trung theo dõi.

Khi đã quyết định tên bài báo cáo, tác giả cần phải xem xét lại cẩn thận. Cần phải xóa bỏ những từ rườm rà vì những từ này có thể làm cho người nhìn khó lĩnh hội vấn đề. Tránh dùng tên bài mang tính quá kỹ thuật. Người dự hội nghị thỉnh thoảng nghe các bài nói chuyện ngoài lĩnh vực chuyên môn của họ, nhưng họ có thể nghĩ rằng những kiến thức và kết quả trong bài nói chuyện có thể giúp ích cho lĩnh vực nghiên cứu của họ. Do đó, tác giả nên cố gắng đặt tên bài sao cho khán giả cảm thấy gần gũi, không quá chi tiết kỹ thuật. Cuối cùng, tên của một báo cáo khoa học không thể để sai chính tả hay ngữ pháp, dù đó là báo cáo bảo vệ luận án hay báo cáo trong hội nghị khoa học.

* Bản chiếu đầu tiên

Bài báo cáo dĩ nhiên bắt đầu bằng bản chiếu đầu tiên. Thông thường trong bản chiếu đầu tiên ít nhất có 2 thông tin quan trọng là:

1. Tên bài báo cáo (thường viết với cỡ chữ 40 trở lên để khán giả dễ đọc);
2. Tên tác giả và nơi làm việc;

Ngoài ra, một số báo cáo còn có thể cần cung cấp thêm các thông tin như:

3. Tên và ngày hội nghị;
4. Danh sách đồng tác giả (nếu có);
5. Tên và logo của cơ quan đào tạo/nghiên cứu;
6. Tên của người hướng dẫn khoa học (nếu là NCS);
7. Cảm tạ;
8. Cơ quan tài trợ;
9. Hình ảnh nền.

Thông tin thứ 3 có khi cần thiết, vì nó cho thấy báo cáo viên có đầu tư thời gian để soạn tài liệu cho hội nghị. Thông tin 5-7 có khi không cần thiết vì có thể nói phần lớn người nghe chẳng cần biết tên hay logo của cơ quan hay tên của thầy cô, hay cảm tạ. Tuy nhiên, những thông tin này có khi làm cho thầy cô và đồng nghiệp hài lòng. Thông tin 8 (tên cơ quan tài trợ, nếu có) cũng có khi quan trọng. Có hội nghị yêu cầu tác giả phải nói rõ cơ quan tài trợ để những người tham dự biết được. Thông tin 9 (hình nền) có thể làm cho bản chiếu hấp dẫn hơn, nhưng cần phải chú ý đến hình ảnh. Thông thường, hình ảnh nền là những yếu tố của công trình hay cơ quan nghiên cứu.

Cần lưu ý rằng bản chiếu đầu tiên không nên cung cấp quá nhiều thông tin. Nhiều thông tin quá rất dễ làm cho khán giả bị sao lãng. Tùy theo loại hội nghị, đối tượng khán giả và tùy theo yêu cầu của ban tổ chức, thông thường chỉ cần tên bài báo cáo và tên tác giả là đủ.

* Soạn các bản chiếu

Cách trình bày bản chiếu có vai trò rất quan trọng trong việc chuyển tải thông tin cho người nghe. Người báo cáo khoa học phải tự mình soạn các bản chiếu theo phong cách của mình và phải am hiểu mình định nói gì. Mục tiêu là giúp cho người nghe lĩnh hội thông tin nhanh và chú ý theo dõi bài báo cáo của mình.

Sau đây là vài hướng dẫn cho cách soạn bản chiếu hiệu quả:

- Mỗi bản chiếu cần có một tựa đề. Tựa đề trên mỗi bản chiếu cũng giống như bảng chỉ đường, dẫn dắt câu chuyện một cách logic và lí thú. Do đó, tác giả cần phải suy nghĩ cách đặt tựa đề cho mỗi bản chiếu sao cho đơn giản nhưng đủ để khán giả biết mình đang ở đâu trong câu chuyện.

- Mỗi bản chiếu chỉ nên trình bày một ý tưởng. Cũng như trong một đoạn văn (paragraph), không nên có hơn một ý tưởng trên một bản chiếu. Do đó, tất cả các ý, bảng biểu trong bản chiếu chỉ nên dùng để hỗ trợ cho một ý tưởng chính. Ý tưởng của bản chiếu có thể thể hiện qua tựa đề của bản chiếu. Nếu tựa đề bản chiếu không chuyên tải được ý tưởng một cách nhanh chóng, thì người báo cáo sẽ phải tốn thì giờ giải thích và có thể làm loãng hay làm cho khán giả sao nhãng vấn đề.

- Trình bày bản chiếu ngắn gọn. Nếu bản chiếu có quá nhiều chữ thì khán giả sẽ đọc chứ không nghe. Nhưng người báo cáo cần khán giả phải nghe hơn là đọc (vì họ có thể đọc bài báo hay báo cáo khoa học chi tiết hơn). Do đó, nguyên tắc chung là bản chiếu càng ít chữ càng tốt. Mỗi bản chiếu, nếu chỉ có chữ, thì nên tuân thủ theo công thức “n x n”, có nghĩa là nếu quyết định mỗi bản chiếu có 5 dòng chữ thì mỗi dòng chỉ nên có 5 chữ. Một bản chiếu không nên có quá 6 dòng chữ ($n < 7$). Cách viết bản chiếu tốt nhất là cách viết tốc ký. Đó là cách viết không tuân theo văn phạm chính thống, không cần phải có một câu văn hoàn chỉnh, cố gắng viết ngắn, dùng chủ yếu là các từ khoá, bỏ những từ không cần thiết. Không được phép chép từ văn bản soạn thảo thường (text) sang bản chiếu PowerPoint vì như vậy sẽ làm phá vỡ nguyên tắc “kiệm từ”. Đó là một sai lầm phổ biến vì người chuẩn bị thường sợ bị “thiếu” khi chuẩn bị bản chiếu.

- Không dùng quá nhiều hiệu ứng và dấu đầu dòng. Dùng nhiều hiệu ứng chuyển tông (animation) để chuyển bản chiếu hay chuyển ý trong từng bản chiếu sẽ làm cho khán giả mất tập trung vào thông tin chính của báo cáo. Hơn nữa, nó làm cho khán giả cảm thấy đang xem “biểu diễn” hơn là nghe báo cáo khoa học. Do đó nếu dùng hiệu ứng này khi soạn bản chiếu thì cần hạn chế ở những hình thức đơn giản nhất, không được để khi trình bày bản chiếu và/hay chữ chạy lòng vòng hay nhảy múa trước khi dừng. Các dấu đầu dòng (bullet) cũng thường hay được sử dụng trong các bài báo cáo bằng PowerPoint, nhưng cần phải cân nhắc không nên dùng quá nhiều. Nguyên tắc là không lặp lại cùng những từ/cụm từ giống nhau ở trong các dấu đầu dòng liên tiếp.

- Dùng biểu đồ và hình ảnh phù hợp. Biểu đồ dễ gây ấn tượng hơn là con số hay chữ. Phần lớn báo cáo khoa học bằng PowerPoint thường có nhiều biểu đồ và hình ảnh. Việc sử dụng linh hoạt những hình ảnh, kể cả hình ảnh động, sẽ có tác dụng thu hút sự chú ý của khán giả. Hiệu ứng trợ giúp việc xuất hiện của dòng chữ và con số sẽ ngăn ngừa sự sao nhãng của khán giả. Chúng cũng giúp giải thích dữ liệu một cách hiệu quả. Tuy nhiên, chỉ nên sử dụng những hình ảnh đơn giản và không nên quá lạm dụng, nhất là những hình ảnh màu mè và hình ảnh động, nếu không sẽ làm cho khán giả mất tập trung vào nội dung thuyết trình, giảm sự trang trọng của báo cáo khoa học.

- Dùng font chữ dễ đọc. Có hai nhóm font chữ chính: nhóm chữ không có chân và nhóm có chân. Nhóm không có chân bao gồm Arial, Comic Sans, Papyrus, ... Nhóm font chữ có chân bao gồm Times New Roman, Courier, Script, ... Nhiều nghiên cứu tâm lý chỉ ra rằng font chữ không chân thường dễ đọc. Người đọc tốn ít thời gian để đọc các font chữ như Arial hơn là Times hay Times New Roman. Chính vì thế nên dùng font chữ Arial hay các font tương tự.

- Không dùng cỡ chữ quá nhỏ. Nên dùng cỡ (size) từ 18 trở lên. Nếu dùng font chữ với cỡ <18 khán giả sẽ khó đọc, nhất là trong các hội trường rộng. Riêng tên bài báo cáo, cỡ font chữ nên từ 40 đến 50. Tuy nhiên, trong trường hợp phải trình bày tài liệu tham khảo thì font size khoảng 12-14 có thể chấp nhận được.

- Không dùng chữ viết hoa trong phần nội dung của bản chiếu. Chữ viết hoa khó đọc và khó theo dõi. Tuy nhiên, có thể viết nghiêng hay tô đậm, nhưng đừng nên lạm dụng những cách viết này. Chỉ dùng gạch chân khi cần nhấn mạnh một điều gì quan trọng; nếu không thì nên tránh.

- Chọn màu thích hợp. Màu đỏ và màu cam là màu “cao năng” nhưng rất khó tập trung. Màu xanh lá cây, xanh nước biển, và nâu là những màu “ngọt dịu”, nhưng khó gây chú ý. Màu đỏ và xanh lá cây có thể khó thấy đối với những người với hội chứng mù màu. Cách chọn màu còn tùy vào bối cảnh và môi trường. Cũng cần phân biệt màu chữ (text) và màu nền (background). Nguyên tắc chọn màu cho bản chiếu:

+ Nếu hội trường nhỏ: chọn chữ màu tối trên nền sáng, như chữ màu đen hay màu xanh đậm trên nền trắng;

+ Nếu hội trường rộng lớn: chọn chữ sáng trên nền tối, như chữ màu trắng hay vàng trên nền xanh đậm.

+ Tránh làm bản chiếu có chữ màu xanh lá cây và màu nền đỏ hay chữ màu đỏ trên nền màu xanh lá cây vì rất nhiều người bị mù màu với sự kết hợp này. Nói chung tránh chọn màu nền đỏ vì đây là loại màu “cao năng” dễ làm cho mắt bị mệt và khó theo dõi.

+ Hạn chế dùng các bản chiếu nhiều màu mè, “đồng bóng”, không nên có 3 màu trở lên.

3.2.2. Thuyết trình báo cáo

a. Bố cục trình bày báo cáo

Nguyên tắc bố cục bài trình bày cũng theo trình tự chuẩn bị các bản chiếu. Vì vậy, sau khi giới thiệu về báo cáo và những nội dung sẽ trình bày trong vòng một vài phút, người báo cáo cần đi thẳng vào trình bày lần lượt từng nội dung, sau đó tóm tắt lại những gì đã được trình bày trong báo cáo. Cuối cùng không quên nói lời cảm ơn khán giả đã chú ý theo dõi và mời khán giả đặt câu hỏi và/hay bình luận về bài báo cáo của mình.

Nếu người báo cáo làm cho người nghe cảm thấy nhàm chán ngay ở những phút đầu thì việc lấy lại hứng thú cho họ là gần như không thể. Hơn nữa, khi trình bày một báo cáo khoa học không nên tập trung quá nhiều vào các chi tiết hay một vấn đề nào đó, chỉ nên nói vừa đủ, sao cho toát lên được ý chính của vấn đề cần nói. Không nên làm cho người nghe phải thu nhận quá nhiều thông tin, lúc đó vấn đề sẽ trở nên nhàm chán và không mang lại hiệu quả như mong muốn. Không áp dụng một công thức, một cách

trình bày cứng nhắc mà tập trung sâu hơn vào một số khía cạnh, lướt hoặc bỏ qua một số khía cạnh khác. Trong quá trình báo cáo, nếu vì lý do thời gian thì nên tập trung vào một vài phần và bỏ qua các phần khác. Như thế sẽ tốt hơn là đưa ra tất cả ý chính một cách sơ lược mà không đi sâu vào các chi tiết hoặc trình bày vượt quá thời gian cho phép.

b. Phong thái và cử chỉ

- Thể hiện sự tự nhiên và tự tin. Cố gắng giữ điệu bộ và thuyết trình một cách tự nhiên như đang trò chuyện với khán giả. Một dáng điệu và sự di chuyển tốt sẽ truyền tải được sự tự tin, chuyên nghiệp và đáng tin cậy. Dùng cử chỉ để nhấn mạnh các điểm chính và thu hút sự chú ý nơi khán giả, nhưng tránh làm các động tác không tự nhiên, lạ thường hay những động tác quá mạnh (đặc biệt là thói quen vung tay khi nói). Tránh nói một cách đều đều như trả bài bắt buộc hay chỉ nhìn và đọc bản chiếu đã chuẩn bị sẵn.

- Chọn vị trí và tư thế đi đứng phù hợp. Chọn đứng ở vị trí sao cho không che tầm nhìn của khán giả lên màn hình. Không đứng quay lưng lại khán giả và nói với... màn hình (!). Không cúi đầu xuống nhìn các phiếu ghi chú hay bản in và nói với... mặt bàn. Không “đứng như trời trồng” tại một chỗ, nhưng cũng đừng quá nhảy múa lăng xăng.

- Thể hiện sự tích cực và nhiệt tình. Người báo cáo phải thể hiện quan điểm rõ ràng và tích cực, thể hiện sự đầu tư, yêu thích chủ đề đang nói thông qua giọng nói và các biểu cảm trên nét mặt. Không xin lỗi trước về sự thiếu chuẩn bị hay về các khiếm khuyết của bài thuyết trình, vì điều đó chỉ gây sự chú ý của người nghe vào các khiếm khuyết đó mà thôi.

- Giao tiếp thân thiện với khán giả. Duy trì sự giao tiếp bằng mắt với khán giả để tăng sự tin cậy, tăng sự thích thú, tập trung nơi khán giả, và nhờ vậy cũng có thể nhận ra được sự phản hồi ngầm từ khán giả đối với bài thuyết trình của mình. Trong khi nói, nên hướng ánh mắt về phía người nghe và tốt nhất là nhìn từng người và lướt qua khắp phòng. Làm như vậy người thuyết trình sẽ tự tin hơn và người nghe cũng cảm thấy dễ chịu hơn vì họ cảm thấy được mời đến tham dự một cuộc nói chuyện, một cuộc bàn luận chứ không phải tới chỉ để nghe. Cũng vì vậy mà không được giảng như lên lớp mà dùng từ sao cho người nghe thấy thoải mái. Tránh những điệu bộ thiếu thân thiện và không tôn trọng khán giả như bỏ tay vào túi quần, chỉ trở vào khán giả, búng tay, khoanh tay ngang ngực, ...

- Dùng bút chỉ (pointer) để hướng dẫn sự chú ý của khán giả. Để giải thích các chi tiết quan trọng trên màn hình nên dùng pointer để chỉ, không nên rê chuột để chỉ, không nói từ đầu đến cuối mà không chỉ vào đâu cả. Pointer giúp cho khán giả dễ theo dõi người báo cáo đang nói về vấn đề gì và ý nghĩa của nó ra sao. Đi cùng với việc dùng pointer, người báo cáo còn phải nói, nhưng nói ngắn gọn sao cho khán giả lĩnh hội được ý nghĩa của các con số quan trọng hay một hình ảnh/biểu đồ. Cần mô tả vị trí cần chú ý nào thì chỉ pointer vào đúng chỗ đó một cách dứt khoát, không khua bút để điểm sáng chạy lòng vòng vô định trên màn hình.

c. Ngôn ngữ và phát âm

- Thuyết trình rõ ràng. Cố gắng nói lớn, phát âm rõ tiếng, không nói quá nhanh, không nuốt chữ hay gần giọng. Sử dụng ngôn ngữ phù hợp với đối tượng khán giả đang nghe. Không nói ấp úng, không dùng từ đệm để đệm câu hoặc chuyển ý.

- Sử dụng tốt các câu dẫn ý. Người báo cáo cần dùng những cụm từ thích hợp để báo cho khán giả biết mình sắp chuyển sang một điểm khác, hay muốn nhấn mạnh một điểm nào đó. Dấu hiệu chuyển ý trong bài thuyết trình cũng gần giống như trong một bài viết. Chúng giúp người nói chuyển ý và đảm bảo cho người đọc/nghe theo dõi được tốt hơn. Thường trên bản chiếu không có dấu hiệu chuyển ý mà người báo cáo phải dùng lời nói để thay thế. Có một số cụm từ mà người báo cáo có thể dùng luân phiên (không nên dùng một cụm từ suốt thời gian trình bày), những cụm từ thường dùng tùy thuộc vào tình huống. Dưới đây là một số tình huống và cách nói để tham khảo:

+ Để mở đầu, nên theo trình tự sau: (a) Chào mở đầu, (b) Nói lời cảm ơn Chủ tọa¹ đã giới thiệu, (c) Nói lời xã giao với (khen ngợi) người báo cáo trước mình để gây cảm tình (nếu cần), (d) Tự giới thiệu mình (nếu chủ tọa chưa giới thiệu) và giới thiệu chủ đề của bài báo cáo với khán giả.

Ví dụ: (a) Kính chào quý vị! (b) Trước tiên tôi xin được cảm ơn GS. TSKH. Cù Xuân Dần đã có lời giới thiệu và cho phép tôi được vinh dự trình bày báo cáo trước Hội nghị khoa học quan trọng này; (c) Thật “không may” cho tôi là báo cáo của tôi lại đi sau một báo cáo rất tuyệt vời của GS. TS. Nguyễn Văn Tuấn, một nhà khoa học quá nổi tiếng; (d) Tuy nhiên, tôi sẽ cố gắng trình bày với quý vị Báo cáo tóm tắt kết quả thực hiện đề tài “Kiềm hóa rơm lúa tươi làm thức ăn cho trâu bò”.

+ Để thêm thông tin có cùng ý nghĩa với thông tin mới nói xong, có thể dùng những từ sau đây: *thêm vào đó, ngoài ra, hơn nữa, tương tự, thứ nhất, thứ hai, thứ ba, cuối cùng là,*

+ Để cho ví dụ minh họa: *chẳng hạn như, tôi xin đưa ra một ví dụ cụ thể như sau, hãy xem ví dụ cụ thể sau đây, ...*

+ Nhấn mạnh hay giải thích thêm những gì đã nói: *như đã trình bày, nói một cách khác là, điều đó có nghĩa là, ...*

+ Nhấn mạnh kết luận như là một hệ quả của phần trình bày trước đó: *Do vậy mà, hậu quả là, kết cục là, ...*

+ Tóm lược những điểm đã trình bày: *nói tóm lại, kết luận lại là, v.v.*

+ Có một cách chuyển ý khá hữu hiệu là đặt câu hỏi. Mình tự đặt câu hỏi và ... tự trả lời. Với cách này, khán giả sẽ chú ý, vì họ chờ câu trả lời, và đó cũng là một cách không cho họ ngủ!

¹ Nhiệm vụ chính của Chủ tọa trong hội nghị khoa học là giới thiệu người báo cáo (diễn giả), điều hành phiên họp diễn ra theo đúng thời gian đã ấn định, khơi mào thảo luận cuối mỗi báo cáo (kể cả đặt câu hỏi khi không có khán giả nào hỏi). Chủ tọa không có nhiệm vụ tóm tắt bài báo cáo.

d. Thời lượng báo cáo

Nguyên tắc chung là không được trình bày quá thời lượng cho phép. Nói quá dài, người báo cáo sẽ bị chủ tọa cắt ngang, hoặc làm cho những người còn lại phải nói ngắn gọn hơn, hoặc làm cho buổi hội thảo kết thúc muộn. Sự chú ý của người nghe là có giới hạn và những gì người báo cáo cố nói thêm sẽ không hữu ích.

Những người có kinh nghiệm thường đặt thời gian trình bày cho mỗi bản chiếu; thông thường là 1 phút/bản chiếu. Khi đó cần ngừng đặt một điểm mốc trong mỗi bản chiếu và mỗi lần nói tới mốc đó lại khéo léo kiểm tra thời gian một lần (không nên để khán giả biết mình xem đồng hồ). Mặt khác, họ cũng thường chuẩn bị dự phòng một vài bản chiếu ở cuối đề phòng trường hợp bị “cháy” thì sẽ dùng tới. Tuy nhiên, tốt nhất là sắp xếp thời gian nói sao cho vừa đủ, không phải dùng tới những bản chiếu như vậy.

Cố gắng tránh những sai lầm làm cho người nghe cảm thấy chán như: báo cáo viên đọc bản chiếu; chữ quá nhỏ, khó đọc; câu dài, không có dấu đầu dòng; màu khó nhìn; chữ chạy lòng vòng, hoạt cảnh nhiều; dùng âm thanh đệm vào chữ; hình minh họa quá phức tạp. Chuẩn bị kỹ càng trước khi trình bày, kể cả luyện tập thuyết trình trước khán giả, cũng là một cách để tỏ lòng tôn trọng khán giả và thể hiện tính chuyên nghiệp của mình.

Tóm lại, báo cáo khoa học là để trình bày kết quả nghiên cứu, người báo cáo phải đảm bảo rằng những bản chiếu là do mình soạn ra để trình bày kết quả nghiên cứu của chính mình. Khi thuyết trình báo cáo, nhà khoa học cũng phải tỏ ra am hiểu, biết cách diễn giải một cách chủ động, tự nhiên và tự tin. Nói một cách khác là nhà khoa học phải làm cho con số của mình “biết nói” hay “nói hộ” các con số đó.

3.2.3. Trả lời chất vấn sau thuyết trình

Sau một bài thuyết trình báo cáo khoa học lúc nào cũng có phần vấn đáp (Q&A). Đây là phần quan trọng không kém phần nội dung. Sau đây là một vài nguyên tắc theo gợi ý của Giáo sư Nguyễn Văn Tuấn (2012):

- Mời khán giả đặt câu hỏi hay bình luận. Sau khi trình bày xong báo cáo và nói lời cảm ơn khán giả, người báo cáo cần chủ động mời khán giả đặt câu hỏi và/hay bình luận về bài báo cáo của mình.

- Lắng nghe và hiểu rõ câu hỏi. Tập trung chú ý để nghe rõ các câu hỏi của khán giả và phải đảm bảo hiểu đúng ý câu hỏi trước khi trả lời. Nếu chưa nghe rõ phải hỏi lại cho chắc chắn, nhất là khi có người hỏi nhiều câu hỏi một lúc hay trong hội thảo bằng ngôn ngữ nước ngoài. Không được trả lời khi chưa biết được người ta muốn hỏi gì.

- Lịch sự với khán giả. Khi khán giả đang hỏi hay bình luận, không được ngắt lời họ, mà phải chờ cho họ nói xong mình mới trả lời. Nếu có những câu hỏi “kém thông minh” thì người báo cáo cần phản ứng “lịch sự” bằng cách xin phép “không trả lời ở đây”. Tuyệt đối không lên lớp người hỏi mình dù cho câu hỏi có vô duyên hay kém cỏi đến đâu. Không bao giờ nổi nóng tranh cãi khi có người hỏi “xác” hay muốn lên lớp

mình mà phải bình tĩnh tìm cách khéo léo “xua đuổi” câu hỏi/bài giảng của họ. Ví dụ, có thể nói “tôi xin phép không trả lời câu hỏi này ở đây vì thời gian không cho phép, nhưng tôi đã đề cập đến vấn đề này trong bài viết đầy đủ của tôi rồi”. Đó là cách nói gián tiếp “anh hãy về đọc lại bài của tôi”.

- Trả lời ngắn gọn, rõ ràng. Thông thường chỉ có khoảng 5 phút cho hỏi và trả lời sau mỗi báo cáo. Do vậy trả lời phải ngắn gọn, đi thẳng vào vấn đề.

- Biết xử lý câu hỏi khó. Nếu câu hỏi khó quá hay ngoài khả năng trả lời của mình thì người báo cáo có thể né tránh một cách lịch sự. Có một cách khác, nhất là đối với nghiên cứu sinh, là “đá bóng” sang nhờ thầy của mình hay một “cây đa, cây đề” nào đó đang có mặt.

- Chuẩn bị dự phòng. Nên chuẩn bị trước một số bản chiếu dự phòng dành riêng cho những vấn đề không trình bày trong bài thuyết trình nhưng có thể được hỏi đến.

3.3. BÀI BÁO KHOA HỌC

Bài báo khoa học (scientific paper) là một bài báo có nội dung khoa học được công bố trên một tạp chí khoa học (scientific journal) đã qua phản biện khoa học (peer-review). Có một số dạng bài báo khoa học khác nhau: bài báo gốc (original paper), bài tổng quan (review paper), bài báo ngắn (short communication)... Những bài đăng trong các kỷ yếu hội thảo khoa học (proceedings papers) nói chung chưa được coi là bài báo khoa học. Đối với NCS thì một yêu cầu bắt buộc là phải công bố được kết quả nghiên cứu của mình dưới dạng các bài báo khoa học gốc. Do vậy, mục này chỉ đề cập đến bài báo gốc công bố kết quả nghiên cứu chăn nuôi.

3.3.1. Cấu trúc bài báo khoa học

Tùy theo yêu cầu của từng tạp chí, nhưng nhìn chung một bài báo gốc công bố kết quả nghiên cứu chăn nuôi thường có những phần sau:

1. Tên bài báo (Title)
2. Tác giả (Authors)
3. Tóm tắt (Abstract)
4. Giới thiệu/Đặt vấn đề (Introduction)
5. Vật liệu và Phương pháp (Materials and Methods)
6. Kết quả (Results)
7. Thảo luận (Discussion)
8. Kết luận (Conclusions)
9. Cảm tạ (Acknowledgements)
10. Tài liệu tham khảo (References)

Tuy nhiên, cũng có những tạp chí quy định cấu trúc khác với cấu trúc trên, ví dụ *Kết quả* và *Thảo luận* đi liền với nhau thành phần *Kết quả và Thảo luận* hay ghép phần *Kết luận* vào phần *Thảo luận*. Do đó, tác giả cần phải tìm hiểu kỹ quy định của Tạp chí định gửi bài đăng để cấu trúc bài báo cho phù hợp.

3.3.2. Chuẩn bị trước khi viết

Trước khi viết một bài báo cụ thể tác giả cần có bước chuẩn bị với các nội dung sau: (1) Kiểm tra ý tưởng gốc, (2) Quyết định dạng bài báo, (3) Xác định người đọc là ai, (4) Chọn tạp chí để đăng và (5) Đọc hướng dẫn tác giả của tạp chí.

(1) Kiểm tra ý tưởng gốc

Trước tiên tác giả cần kiểm tra lại ý tưởng gốc cho bài báo định viết thông qua việc trả lời các câu hỏi sau đây:

- Công trình nghiên cứu có gì mới và hay không?
- Có thách thức gì trong công trình nghiên cứu không?
- Công trình nghiên cứu có trực tiếp liên quan đến một vấn đề nóng hiện tại không?
- Có thể đưa ra giải pháp cho vấn đề đó không?
- Nếu tất cả câu trả lời là “có” thì có thể bắt đầu viết bản thảo bài báo.

(2) Quyết định dạng bài báo

Tuỳ theo tính chất của thông tin, tác giả có thể chọn một trong số các dạng bài báo khoa học khác nhau sau đây để viết bản thảo:

- Bài báo gốc đầy đủ (original paper): là loại bài báo quan trọng nhất, công bố lần đầu tiên kết quả toàn phần hay từng phần quan trọng đã kết thúc và có ý nghĩa của công trình nghiên cứu.

- Bài báo ngắn/thông báo nhanh (short communication): là bài công bố nhanh và sớm những kết quả ban đầu có ý nghĩa. Bài dạng này ngắn hơn nhiều so với bài báo đầy đủ (thường mỗi tạp chí có giới hạn cụ thể).

- Bài tổng quan (Review paper/perspectives): là bài viết tóm tắt những nghiên cứu gần đây về một chủ đề cụ thể, nêu bật những điểm quan trọng đã được công bố mà không giới thiệu những thông tin mới. Thông thường bài dạng này được viết theo đặt hàng của Tạp chí.

Tác giả tự đánh giá công trình của mình xem đã đủ cho một bài báo gốc đầy đủ chưa hay kết quả rất thú vị cần thông báo càng sớm càng tốt. Nghiên cứu sinh nên hỏi thầy hay đồng nghiệp để xin lời khuyên. Đôi lúc người ngoài thấy rõ vấn đề hơn.

(3) Xác định đối tượng người đọc

Một điều rất quan trọng là tác giả phải xác định được đối tượng người đọc cho bài báo định viết. Một bài báo viết theo ngôn ngữ toán học trừu tượng không thể phù hợp với một kỹ sư thực hành muốn tìm cái gì đó có thể ứng dụng được ngay. Trái lại, đối với một hội nghị khoa học thì một bài báo viết theo kiểu thực hành sẽ không được đánh giá cao.

Bản thân *Tạp chí Khoa học Kỹ thuật Chăn nuôi* của Hội Chăn nuôi Việt Nam cũng phân ra hai loại số xuất bản khác nhau tùy theo đối tượng người đọc: *Khoa học - Công nghệ* (dành chủ yếu cho giới nghiên cứu/học thuật) và *Sản xuất - Thị trường* (dành chủ yếu cho giới sản xuất/kinh doanh). Do đó bài viết cho hai loại số xuất bản này cũng cần khác nhau phù hợp với tính chất của người đọc.

(4) Chọn tạp chí để đăng

Tác giả cần tìm hiểu tất các tạp chí có thể đăng để xem: (1) mục tiêu và phạm vi, (2) loại bài báo, (3) loại người đọc, và (4) các chủ đề nóng mà Tạp chí quan tâm (xem qua các tóm tắt những bài báo gần đây). Đồng thời cũng nên xin lời khuyên của thầy và đồng nghiệp nên gửi đăng ở tạp chí nào thì phù hợp nhất. Các bài báo được trích dẫn trong bài báo thường là chỉ dẫn tốt cho tác giả tìm đúng tạp chí để đăng. Không được gửi bản thảo đến nhiều tạp chí, chỉ gửi một lần. Chuẩn mực quốc tế cấm gửi bài tới nhiều tạp chí cùng một lúc và các ban biên tập cũng sẽ dễ dàng phát hiện ra nếu tác giả làm điều đó.

(5) Đọc hướng dẫn tác giả của Tạp chí

Tác giả phải áp dụng đúng hướng dẫn của Tạp chí định đăng bài để viết bản thảo, ngay từ bản thảo đầu tiên (trình bày, trích dẫn tài liệu tham khảo, cách thể hiện bảng biểu...). Điều đó giúp tiết kiệm thời gian cho cả tác giả và Ban biên tập. Nghiên cứu sinh cũng đừng nghĩ rằng cứ viết đại đi rồi thầy sửa cho. Thầy chỉ “hướng dẫn” chứ không làm thay, không làm mất thời gian quý giá của thầy để sửa lỗi viết và thể thức bài báo.

Ví dụ: Khi định gửi bài đăng trên Tạp chí Khoa học và Phát triển của Học viện Nông nghiệp Việt Nam tác giả phải tuân theo hướng dẫn viết nội dung của các phần như sau (trích):

“... Tiêu đề bài báo phải bao hàm nội dung bài viết, ngắn gọn súc tích, đủ nghĩa, sử dụng thuật ngữ khoa học đại chúng và thể hiện được từ khoá.

Tóm tắt tiếng Anh nêu được mục đích, phương pháp nghiên cứu chính, kết quả nghiên cứu và kết luận chủ yếu, không quá 250 từ. Từ khoá khoảng 3-6 từ bằng cả tiếng Việt và tiếng Anh.

Nội dung của các phần: (1) Phần mở đầu: cung cấp bối cảnh/cơ sở của vấn đề, tổng quan các tài liệu liên quan đến vấn đề nghiên cứu, logic dẫn đến việc nghiên cứu và trình bày mục đích nghiên cứu. (2) Phần vật liệu và phương pháp nghiên cứu: mô tả các vật liệu dùng cho nghiên cứu, phương pháp nghiên cứu, địa điểm và thời gian bố trí các thí nghiệm, các chỉ tiêu theo dõi và phương pháp xử lý số liệu. (3) Phần kết quả và thảo luận: trình bày các kết quả thu được theo trình tự logic, không nhắc lại số liệu đã có trong các bảng biểu; thảo luận giúp cho việc diễn giải các kết quả nghiên cứu, khám phá những mối quan hệ với các nghiên cứu trước đó thông qua tài liệu tham khảo, giải thích được sự quan trọng, cũng như tính hợp lý của kết quả nghiên cứu...”

3.3.3. Cách viết các phần của bài báo

a. Tên bài báo (Title)

Tên bài báo nhằm thông báo bài báo đại khái nói về cái gì. Một cái tên tốt sử dụng ít từ ngữ nhất nhưng đủ để mô tả nội dung của bài báo. Tên bài báo quan trọng nên phải được lựa chọn cẩn thận vì sẽ được hàng nghìn người đọc trong khi chỉ có một số ít người sẽ đọc toàn bài. Việc người đọc tìm được bài báo phụ thuộc vào sự chính xác của tên bài, bài báo có tên không phù hợp sẽ dễ bị lãng quên. Thường có 3 cách chính để viết tên bài báo:

- Một cụm từ mô tả/trần thuật. Ví dụ: “*Ảnh hưởng của kiểm hóa bằng urê đến tỷ lệ tiêu hóa của rơm lúa*”. Đây là cách viết phổ biến nhất, đặc biệt là khi mô tả mối quan hệ nhân - quả.

- Một tuyên bố mang tính tuyên ngôn hay kết luận vấn đề. Ví dụ: “*Kiểm hóa rơm bằng urê làm tăng tỷ lệ tiêu hóa của rơm*”. Cách viết này không được khuyến khích vì trong khoa học không có cái gì có thể xác định một cách chắc chắn như một chân lý.

- Một câu hỏi. Ví dụ: “*Loại bò sữa nào nên nuôi ở Việt Nam?*” Cách viết này cũng không được dùng phổ biến trong các bài báo khoa học.

Sau đây là một số lời khuyên để có một tên bài báo tốt:

- Không nên quá ngắn (dưới 9 từ), nhưng cũng không nên quá dài (quá 20 từ)
- Chứa đựng càng nhiều các từ khóa (keyword) của bài báo càng tốt
- Phản ánh nội dung chính của công trình nghiên cứu
- Dùng từ ngắn gọn, súc tích, cụ thể và có thông tin
- Cần có ý nghĩa, không chung chung, mơ hồ
- Cần thể hiện được bản chất của thí nghiệm và các phát hiện mới
- Không chứa các chữ viết tắt, công thức hóa học, tên riêng hay từ lóng
- Hết sức cẩn thận tránh lỗi ngữ pháp và lỗi chính tả

Tên bài báo có vai trò quan trọng vì nó là dòng chữ đầu tiên mà người đọc thấy và sẽ quyết định có nên đọc tiếp hay không. Có thể bắt đầu bằng việc liệt kê những từ khóa quan trọng nhất rồi nghĩ ra một tiêu đề có chứa những từ này. Tiêu đề có thể nêu lên kết luận của bài báo. Tuy là dòng chữ đầu tiên của bài báo nhưng nên cân nhắc và viết chót cuối cùng, sau khi nội dung bài báo đã được soạn thảo xong. Tác giả cần suy nghĩ cẩn thận nhiều lần về tên bài báo trước khi gửi bài cho Tạp chí.

b. Tác giả

Mọi bài báo khoa học đều phải có tên tác giả. Thường thì một thành viên nghiên cứu muốn đứng tên tác giả bài báo phải hội tụ đủ 3 tiêu chuẩn chung sau:

- Có đóng góp quan trọng trong việc hình thành ý tưởng nghiên cứu, xây dựng phương pháp nghiên cứu, thực hiện thí nghiệm, thu thập và quản lý dữ liệu, hay phân tích và diễn giải dữ liệu;

- Đã soạn thảo bài báo hay kiểm tra nội dung tri thức của bài báo;
- Đồng ý với bản thảo cuối cùng để gửi cho Tạp chí.

Trên đây là những tiêu chuẩn có tính nguyên tắc, nhưng trong thực tế còn nhiều tranh luận và không phải tác giả nào cũng hội đủ tất cả các điều kiện đó, thậm chí còn có “tác giả danh dự”. Công trình nghiên cứu khoa học ngày nay là công trình của một tập thể, do đó mỗi bài báo khoa học gốc có tên của nhiều tác giả. Vấn đề là ở chỗ sắp xếp thứ tự tên các tác giả như thế nào. Trên nguyên tắc, trình tự tên các tác giả phải dựa vào mức độ đóng góp của từng tác giả. Người có đóng góp lớn nhất phải là tác giả số 1 (đứng tên đầu tiên). Tuy nhiên, trên thực tế đây là một vấn đề khó, tế nhị và còn có nhiều tranh luận, chưa có sự thống nhất chung, nhất là ở Việt Nam. Khó là ở chỗ không có thước đo nào để đo lường chính xác và khách quan mức độ đóng góp của các tác giả. Đối với các công trình nghiên cứu thuộc đề tài tiến sĩ thì tác giả đứng đầu thường là nghiên cứu sinh, sau đó là người hướng dẫn khoa học, tiếp nữa là những người có một số đóng góp nhất định. Tuy nhiên, cũng có quan điểm cho rằng giáo viên hướng dẫn nên đứng tên cuối cùng trong danh sách tác giả của bài báo để dễ nhận dạng.

Hầu hết các tạp chí đều ghi địa chỉ của các tác giả bài báo. Trong số các tác giả thì có một tác giả liên lạc chính và thường có địa chỉ email kèm theo. Ví dụ:

J. Sci. & Devel. 2014, Vol. 12, No. 5: 675-682	Tạp chí Khoa học và Phát triển 2014, tập 12, số 5: 675-682 www.vnua.edu.vn
ẢNH HƯỞNG CỦA THỜI ĐIỂM THU CẮT ĐẾN NĂNG SUẤT, THÀNH PHẦN HÓA HỌC CỦA HAI DÒNG CAO LƯƠNG (OPV86 VÀ OPV88) VÀ CHẤT LƯỢNG THỨC ĂN Ủ CHUA TỪ CÂY CAO LƯƠNG Nguyễn Thanh Nhân¹, Nguyễn Xuân Trạch^{2*}, Bùi Quang Tuấn², Phạm Văn Cường³ ¹Trường Trung cấp Kinh tế - Kỹ thuật Yên Bái, ²Khoa Chăn nuôi và NTTS, Học viện Nông nghiệp Việt Nam, ³Khoa Nông học, Học viện Nông nghiệp Việt Nam Email*: nxtrach@vnua.edu.vn	

c. Tóm tắt/Tóm lược (Abstract)

Đây thường là phần đầu tiên và có khi là phần duy nhất của bài báo mà độc giả sẽ đọc. Nó cần tóm tắt rõ ràng, súc tích những thông tin chính trong bài báo. Tùy theo quy định của từng tạp chí, thông tin tóm tắt này phải được viết trong vòng 150-300 từ, cần có những thông tin quan trọng như sau:

- Câu hỏi và mục đích nghiên cứu. Thông tin này thường được mô tả bằng 2 câu văn: câu thứ nhất mô tả vấn đề/câu hỏi nghiên cứu mà tác giả quan tâm là gì trong tình trạng tri thức hiện tại; câu thứ hai mô tả mục đích nghiên cứu một cách ngắn gọn và rõ ràng.

- Phương pháp nghiên cứu. Mô tả mô hình thiết kế của công trình nghiên cứu, đối tượng nghiên cứu, chỉ tiêu theo dõi và phương pháp xác định. Những thông tin này có thể viết trong vòng 4-5 câu.

- Kết quả. Trình bày những kết quả chính của nghiên cứu trả lời câu hỏi nghiên cứu đã được đặt ra từ câu văn đầu tiên của bản *Tóm tắt*. Những thông tin này có thể viết trong vòng 5-7 câu.

- Kết luận. Dùng 1-2 câu văn viết về những kết luận (khái quát) rút ra được từ công trình nghiên cứu và ý nghĩa của kết quả nghiên cứu. Đây là thông tin mà người đọc thường quan tâm nhất nên cần chọn câu chữ sao cho thuyết phục và thu hút được sự chú ý của người đọc.

Bản Tóm tắt đóng vai trò rất quan trọng vì nó giúp người đọc biết có nên đọc tiếp hay bỏ qua bài báo. Thông thường cứ khoảng 500 người đọc bản *Tóm tắt* thì chỉ có 1 người đọc bản toàn văn của bài báo. Tác giả thường viết phần này sau khi đã hoàn tất bài báo. Nên viết chỉ trong một đoạn văn (paragraph). Trong bản tóm tắt bằng tiếng Anh (Abstract) viết động từ ở thì quá khứ để chỉ công trình đã thực hiện. Các tập hợp từ dài nên mở ngoặc viết tắt và sau đó dùng từ viết tắt trong phần còn lại của tóm tắt cũng như toàn bài báo. Không nên trích dẫn tài liệu tham khảo trong phần này và càng không nên viết thông tin hay kết luận không có trong bài báo. Phải kiểm tra tính chính xác khi trích viết các số liệu quan trọng từ các bảng biểu trong bài báo.

d. Giới thiệu/ Đặt vấn đề (Introduction)

Phần Giới thiệu hay Đặt vấn đề là phần mở đầu của bài báo nhằm giới thiệu và thu hút sự chú ý của người đọc. Điều quan trọng nhất là làm sao để người đọc sau khi đọc xong phần này thấy được tầm quan trọng của vấn đề nghiên cứu, hiểu được tại sao tại sao tác giả lại làm nghiên cứu đó và quyết định đọc tiếp các phần sau của bài báo. Vì thế tác giả cần thể hiện được rằng công trình nghiên cứu có tầm quan trọng về mặt khoa học, câu hỏi nghiên cứu nhằm vào lỗ hổng kiến thức (knowledge gap) hiện thời và nghiên cứu hiện tại nhằm trả lời câu hỏi đó, tức là cho ra kết quả mới, giúp giải quyết một phần của vấn đề lớn.

Có nhiều cách tiếp cận vấn đề khác nhau, thông thường nhà nghiên cứu chăn nuôi thường áp dụng nguyên tắc “từ tổng quan đến cụ thể”, bắt đầu bằng việc mô tả bức tranh lớn (vấn đề chung), sau đó thu hẹp dần vấn đề để đi đến câu hỏi và mục tiêu nghiên cứu. Tác giả cần trả lời câu hỏi chính là “Tại sao lại làm nghiên cứu này?” thông qua việc cung cấp các thông tin để trả lời các câu hỏi sau đây:

- Vấn đề quan tâm nghiên cứu là gì? Tác giả cần nêu được tầm quan trọng của vấn đề nghiên cứu, đó có phải là vấn đề nóng trong lĩnh vực nghiên cứu này không.

- Những gì đã và chưa được nghiên cứu về vấn đề này? Tác giả tóm lược tổng quan những kết quả nghiên cứu đã được công bố liên quan trực tiếp đến vấn đề đang được quan tâm nghiên cứu, nhưng không nêu những khái niệm cơ bản mà những người trong ngành đều đã biết. Nên tập trung vào những nghiên cứu gần đây và tránh dùng những kiến thức trong giáo trình. Đánh giá có biện luận những điểm mạnh và điểm yếu của các nghiên cứu trước, đâu là lỗ hổng kiến thức hiện tại để dẫn dắt đến câu hỏi nghiên cứu và cách giải quyết vấn đề đó trong nghiên cứu của mình.

- Câu hỏi nghiên cứu là gì và tại sao lại quan trọng? Tác giả cần nêu được câu hỏi về vấn đề nghiên cứu mà chưa được trả lời bằng các công trình nghiên cứu trước đây.

- Giả thuyết muốn kiểm định là gì? Tác giả cần nêu dự kiến kết quả trả lời cho câu hỏi nghiên cứu và cách tiếp cận đã sử dụng để kiểm định giả thuyết nêu ra.

- Mục tiêu của nghiên cứu này là gì? Đây là đoạn văn hay những câu cuối của phần Đặt vấn đề. Mục tiêu nghiên cứu phải nêu rất cụ thể. Tác giả có thể phát biểu giả thuyết nghiên cứu và sau đó có câu kiểu như “Nghiên cứu này được thiết kế nhằm kiểm tra giả thuyết đã nêu nhằm mục tiêu ...”.

Các thông tin trên cần viết một cách logic, kết nối chặt chẽ các ý tưởng với nhau, tạo ra một vòng khép kín từ câu đầu đến câu cuối để cho người đọc hiểu được ý định của tác giả là gì và nghiên cứu này nằm ở đâu trong bức tranh tổng thể của chuyên ngành. Không nên viết phần này quá dài vì như thế sẽ làm cho người đọc sao lãng vấn đề chính và có cảm giác rằng tác giả không có khả năng “tổng hợp” mà chỉ biết sao chép các kiến thức rời rạc. Tuy nhiên, cũng không nên viết quá ngắn vì như thế sẽ làm cho người đọc có cảm nhận rằng tác giả thiếu đầu tư suy nghĩ sâu về vấn đề nghiên cứu, thiếu ý tưởng hay thiếu thông tin.

Chú ý: Đối với vài báo viết bằng tiếng Anh thì trong phần này dùng động từ ở thì hiện tại khi nói về các công trình của người khác đã được công bố, nhưng lại dùng thì quá khứ khi nói về công trình đã nghiên cứu của chính mình.

e. Vật liệu và Phương pháp (Materials and Methods)

Khi viết phần này tác giả phải trả lời được câu hỏi là trong nghiên cứu này tác giả đã làm những (nội dung) gì?, làm như thế nào?, thu thập và phân tích số liệu bằng cách nào? để cho ra kết quả nghiên cứu. Đối với các bài báo nghiên cứu chăn nuôi thông thường phần này được viết theo những tiêu mục sau:

- Đối tượng/Vật liệu nghiên cứu (Research Target/Materials). Thông tin về đối tượng và vật liệu nghiên cứu (loại vật nuôi, loại thức ăn ...) phải được nêu cụ thể, trong đó nói rõ nguồn gốc và đặc điểm của từng loại đối tượng/vật liệu nghiên cứu. Khi mô tả đặc điểm của đối tượng nghiên cứu, cần nêu chi tiết tiêu chuẩn lựa chọn (như độ tuổi, khối lượng...) và có thể cả tiêu chuẩn loại trừ (nếu cần). Đối với hóa chất, nêu chính xác các đặc tính kỹ thuật và nguồn gốc hay phương pháp bảo chế. Tránh dùng tên thương phẩm của các hóa chất mà dùng tên hóa học của các hoạt chất/hóa chất được sử dụng.

- Địa điểm, thời gian và bối cảnh nghiên cứu (Settings). Trình bày thông tin về địa điểm, thời gian mà công trình nghiên cứu đã được thực hiện. Cần mô tả điều kiện tiến hành nghiên cứu như nhiệt độ, ẩm độ, ánh sáng... trong khi tiến hành nghiên cứu.

- Thiết kế thí nghiệm (Experimental design). Nêu rõ mô hình thiết kế thí nghiệm nào đã được sử dụng cho công trình nghiên cứu. Ví dụ, “Thí nghiệm được thiết kế theo mô hình nhân tố 2×3 , trong đó 2 loại lợn được thử nghiệm cho ăn 3 khẩu phần thức ăn khác nhau”. Trong phần này cũng cần phải nói về đơn vị thí nghiệm và số lần lặp lại (replicates).

- Các chỉ tiêu nghiên cứu (Parameters/Measurements). Nêu rõ những chỉ tiêu nào được dùng làm thước đo để đánh giá kết quả nghiên cứu và cách xác định chúng như thế nào. Cần mô tả các phương tiện đo lường các chỉ tiêu kèm theo các thông số và độ tin cậy, độ chính xác của từng phương tiện. Chỉ mô tả những chỉ tiêu có liên quan đến bài báo, chứ không phải mô tả tất cả những chỉ tiêu theo dõi trong công trình nghiên cứu.

- Quy trình nghiên cứu (Procedure). Tóm lược từng bước nghiên cứu, chi tiết cho từng can thiệp (nếu có). Nếu quy trình có liên quan đến chọn mẫu thì cần mô tả cụ thể quy trình chọn mẫu như thế nào, kỹ thuật nào đã được sử dụng để đảm bảo rằng các nhóm/lô thí nghiệm là đồng đều,... Ngoài ra, tác giả cần phải mô tả cẩn thận quy trình nuôi dưỡng động vật thí nghiệm (nếu có), phương tiện kỹ thuật đo lường như tên, phiên bản, nơi sản xuất của các máy móc được sử dụng.

- Phân tích số liệu (Data analysis). Tác giả phải nêu được mô hình thống kê sử dụng để phân tích số liệu, trong đó chỉ rõ biến phụ thuộc và biến độc lập là gì và dạng quan hệ giữa chúng. Không tuyên bố chung chung kiểu như “Số liệu thí nghiệm được phân tích theo phương pháp thống kê sinh vật học”. Điều đó chẳng khác nào nói “Lợn thí nghiệm được nuôi theo phương pháp dinh dưỡng học”. Đôi khi tác giả cũng cần phải lý giải tại sao lại chọn phương pháp phân tích này mà không chọn phương pháp phân tích khác. Tác giả cũng cần nêu rõ là đã dùng phần mềm nào để thực hiện việc phân tích thống kê.

Đối với người phản biện thì *Phương pháp nghiên cứu* là phần quan trọng nhất của bài báo. Một nghiên cứu đã cho thấy nguyên nhân lớn nhất để bài báo bị phản biện từ chối đăng là do những khiếm khuyết về phương pháp nghiên cứu hoặc do mô tả phần này không được đầy đủ (Bảng 3.7). Do đó các phần này phải được viết rõ ràng, chính xác, súc tích để đồng nghiệp có thể làm lặp lại được. Nếu là phương pháp mới, phải nêu *toàn bộ các chi tiết*. Nếu các phương pháp nghiên cứu đã được công bố trước đó trong một tạp chí khoa học thì chỉ cần chỉ dẫn nguồn gốc tài liệu tham khảo rõ ràng cho từng phương pháp. Phương pháp phân tích thống kê cũng vậy, các phương pháp thông dụng thì không cần, nhưng phương pháp mới hay không thông dụng thì cần trích dẫn nguồn gốc tham khảo.

g. Kết quả (Results)

Trong phần Kết quả tác giả phải trả lời cho được câu hỏi là công trình nghiên cứu đã phát hiện được những gì? Những gì không phát hiện được so với mục tiêu ban đầu? Vấn đề quan trọng là kết quả phải được trình bày theo một trật tự logic để lần lượt trả lời các mục đích hay câu hỏi nghiên cứu mà tác giả đã nêu ra trong phần Đặt vấn đề. Do vậy, có thể có những tiêu đề nhỏ trong phần Kết quả để giúp người đọc dễ dàng theo dõi và đối chiếu với phần Phương pháp nghiên cứu và các mục tiêu cụ thể của nghiên cứu.

Kết quả trước hết được trình bày dưới dạng bảng số liệu hay biểu đồ (gọi chung là bảng biểu) và được mô tả bằng lời một cách ngắn gọn trong văn bản (text). Chi tiết về cách trình bày kết quả nghiên cứu bằng bảng biểu đã được đề cập trong Mục 3.1 của

chương này. Để mô tả một bảng biểu tác giả nên đi theo các bước sau: (1) Viết một vài dòng giới thiệu bảng biểu đó nói về vấn đề gì và mục đích của nó; (2) Giải thích ý nghĩa các chỉ tiêu; (3) Mô tả dữ liệu theo các chỉ tiêu; (4) Nhấn mạnh những kết quả và phát hiện quan trọng đáng chú ý.

Mục đích của lời viết trong phần Kết quả là để phân tích và dẫn dắt người đọc nhận biết những phát hiện mới, nhấn mạnh những điểm chính yếu. Đây là việc làm cho các con số “biết nói” hay “nói hộ” các con số có trong bảng biểu để làm cơ sở cho bước thảo luận tiếp theo. Đối với bảng số liệu, so sánh theo chiều ngang (giữa các cột) và/hay theo chiều dọc (giữa các hàng) để chỉ ra những khác nhau có ý nghĩa giữa các giá trị trung bình. Chú ý tìm ra các tương tác có nghĩa (nếu có). Đối với biểu đồ, cần chỉ ra được chiều hướng biến thiên của y theo x, đặc biệt là những điểm thay đổi quan trọng như cực đại, cực tiểu, điểm uốn trên đồ thị. Chú ý, các số liệu so sánh của đồng nghiệp (lấy từ các tài liệu tham khảo) chưa dùng ở đây mà sẽ dùng để thảo luận sau.

Khi mô tả kết quả nghiên cứu nếu viết giá trị của các chỉ tiêu nghiên cứu thì phải viết kèm đơn vị đo lường và kết quả phân tích thống kê (nếu có), đồng thời cũng cần đề cập đến xu hướng và mức độ khác biệt. Ví dụ, “Tăng khối lượng bình quân của dê đực cao hơn dê cái 12g/con/ngày ($P < 0,05$)” hay “Tăng khối lượng bình quân hàng ngày của lợn thí nghiệm ở lô A cao hơn 15% so với lợn ở lô B ($P < 0,01$)”. Tuy nhiên, lúc này chưa đưa ra nhận xét về số liệu như cao hay thấp, tốt hay xấu,... vì những nhận xét như vậy mang tính chủ quan của tác giả (chưa chắc người khác đã đồng ý) và sẽ đưa vào phần Thảo luận ngay sau khi trình bày kết quả hay trong một mục riêng.

Nguyên tắc chung khi viết lời mô tả kết quả là: (1) Súc tích và khách quan; (2) Không viết lại những số liệu đã có trong bảng biểu; (3) Cung cấp thông tin bổ sung cho bảng biểu (có thể là những con số được tính toán thêm từ các số liệu có trong bảng số liệu); (4) Có thông điệp rõ ràng.

Khi trình bày kết quả cần tránh những sai lầm sau:

- Đưa vào những thông tin và dữ liệu lặt vặt không hỗ trợ cho mục tiêu nghiên cứu;
- Trình bày dữ liệu không theo thứ tự hay không ăn khớp với mục tiêu nghiên cứu;
- Số liệu không chính xác hay mâu thuẫn với nhau;
- Lặp lại dữ liệu, lạm dụng bảng biểu, dùng cả bảng số liệu và biểu đồ thể hiện cùng dữ liệu;
- Không có kết quả của những nội dung nghiên cứu đã nêu, hay ngược lại;
- Viết lẫn lộn giữa kết quả với phương pháp (mô tả trong phần trước).
- Trích dẫn tài liệu tham khảo (của người khác) hay thảo luận ngay trong phần trình bày Kết quả.

Phần Kết quả là phần cốt lõi của bài báo nên cần được trình bày rõ ràng và đơn giản vì nó tạo nên phần đóng góp kiến thức mới. Chỉ tóm tắt và minh họa những phát hiện mới theo trình tự và logic mà chưa cần giải thích. Đây là phần mang tính khách

quan nên tác giả chỉ mô tả sự thật (facts), kể cả những số liệu thu được trái với tiên đoán/giả thuyết ban đầu của tác giả. Nếu bài báo/tóm tắt viết bằng tiếng Anh thì động từ viết ở thì quá khứ (past tense) để chỉ những gì *đã* được phát hiện.

h. Thảo luận (Discussion)

Thảo luận mang tính *chủ quan* của tác giả diễn ra sau khi trình bày và mô tả kết quả nghiên cứu. Đây là phần khó viết nhất và là nơi tác giả bổ sung giá trị cho công trình nghiên cứu. Chức năng chính của phần này là giải thích giải thích tầm quan trọng của phát hiện, và chỉ ra những đóng góp vào tri thức cũng như định hướng cho tương lai. Thảo luận quan trọng vì tác giả phải sử dụng kết quả phân tích số liệu, biện minh cho phương pháp và đánh giá có biện luận nghiên cứu của mình để trả lời câu hỏi nghiên cứu, chấp nhận hay bác bỏ giả thuyết đã đề ra. Trong phần này, bên cạnh số liệu của mình nhà khoa học có thể dùng số liệu của đồng nghiệp (lấy từ các tài liệu tham khảo) hoặc là để hỗ trợ cho luận điểm của mình hay để làm thông tin nền cho phần thảo luận. Hơn thế, vì thảo luận là nơi để biện luận, đưa ra những luận điểm, nhận thức mới nên tác giả phải suy nghĩ một cách thận trọng về ý nghĩa của kết quả thu được. Đồng thời trong phần này tác giả cũng cần phân tích, nhận biết những hạn chế và tồn tại của nghiên cứu, lý giải cách giải quyết, cải thiện.

Theo Nguyễn Văn Tuấn (2012) cách viết phần thảo luận có thể theo cấu trúc và trình tự như sau:

(1) *Tóm tắt giả thuyết và những phát hiện chính của nghiên cứu.* Đây thường là một đoạn văn tóm tắt lại những ý chính đã nêu trong phần Đặt vấn đề và phần Kết quả để một lần nữa nhấn mạnh rằng giả thuyết của tác giả đã có được “minh chứng”.

(2) *So sánh với kết quả của các nghiên cứu trước.* Trong đoạn này, tác giả phải phát biểu về những kết quả nhất quán và những khác biệt so với các chứng cứ trước đây. Ngoài việc so sánh để tìm ra sự khác biệt hay giống nhau giữa kết quả của nghiên cứu hiện tại với kết quả của các nghiên cứu trước đây của đồng nghiệp và/hay của chính mình (nếu có), tác giả còn phải lý giải tại sao có sự giống nhau hay khác biệt đó. Khi giải thích sự khác biệt giữa các kết quả nghiên cứu cần chú ý đến sự khác biệt về đối tượng nghiên cứu (tuổi, giới tính...), điều kiện của địa điểm nghiên cứu, phương pháp/công cụ đo lường, phương pháp phân tích, v.v... Nếu không tìm được lý do giải thích thì tác giả không được cố biện luận thiếu căn cứ mà phải thẳng thắn thừa nhận là “chưa giải thích được”.

(3) *Giải thích kết quả và cơ sở của các mối liên hệ.* Giải thích tại sao lại có kết quả như trong nghiên cứu, kết quả đó phù hợp hay trái với giả thuyết nghiên cứu đặt ra lúc đầu. Đoạn này khó viết nhất vì tác giả phải hệ thống hóa nhiều kiến thức hiện có để giải thích những kết quả thu được và mối quan hệ giữa chúng. Tác giả có thể trích dẫn các nghiên cứu khác để hỗ trợ cho giải thích của mình. Nếu không biết hay không rõ cơ chế thì tác giả có thể đề xuất giả thuyết mới để giải thích và đề nghị có các nghiên cứu tiếp theo để kiểm định.

Trong khi thảo luận kết quả nghiên cứu của mình hay so sánh với kết quả của các nghiên cứu trước, tác giả không nên viết theo kiểu lí luận một chiều, không sợ những kết quả “không theo ý muốn”. Trong khoa học, bất cứ một kết quả nào cũng phải được giải thích bằng nhiều góc cạnh. Hơn nữa, bất cứ một công trình nghiên cứu tốt nào cũng có hạn chế, nhiều nghiên cứu cũng cho ra những kết quả ngoài dự kiến. Đó là những kết quả không nhất quán với giả thuyết, những dữ liệu nằm ngoài phạm vi, những quan sát lạ. Tác giả cần phải ghi nhận những quan sát lạ và tìm lời giải thích khả dĩ; nếu không giải thích được thì thành thật thú nhận là “chưa giải thích được” và đề nghị có các nghiên cứu khác để làm sáng tỏ. Đó không phải là yếu kém mà biết đâu chính những kết quả lạ này sẽ là đầu mối dẫn đến những khám phá quan trọng sau này nếu có các nghiên cứu tiếp theo. Do đó, tác giả không nên bỏ qua, mà cần ghi nhận và chú giải cẩn thận.

(4) *Khái quát hóa kết quả nghiên cứu và ý nghĩa của kết quả đó.* Trong đoạn này tác giả cần suy luận về khả năng áp dụng những phát hiện trong nghiên cứu cho bối cảnh lớn hơn, những giả định nào cần có cho việc áp dụng phổ biến kết quả đó, ý nghĩa/tác động về kinh tế - xã hội và môi trường của kết quả nghiên cứu.

(5) *Thảo luận những ưu điểm và hạn chế của công trình nghiên cứu.* Bất cứ một nghiên cứu nào, dù được thiết kế cẩn thận đến đâu, cũng đều có những thế mạnh và những hạn chế của nó. Do đó, tác giả cần phải ghi nhận những điểm yếu, những hạn chế của nghiên cứu. Những điểm mạnh và điểm yếu của công trình nghiên cứu có thể liên quan đến cách thiết kế, đối tượng nghiên cứu, bối cảnh thực hiện nghiên cứu, công cụ đo lường, phương pháp phân tích, v.v... Ghi nhận những khiếm khuyết của nghiên cứu không phải là tín hiệu của sự yếu kém; ngược lại, đó là cách thể hiện sự nghiêm túc và sự thận trọng cần thiết của tác giả. Ngoài việc nêu những hạn chế, tác giả cũng có thể viết về sự ảnh hưởng của những hạn chế đến kết quả nghiên cứu. Chẳng hạn như nếu số lượng mẫu quá ít (có thể do thiếu kinh phí mua gia súc hay điều kiện bố trí thí nghiệm) thì kết quả có thể không đáng tin cậy và cần phải ghi nhận điều này.

i. Kết luận (và Đề nghị)

Đây là đoạn văn ngắn nhưng rất quan trọng vì nó phải chuyển tải được cho người đọc những thông điệp quan trọng từ kết quả của công trình nghiên cứu thông qua suy luận *quy nạp*. Kết luận có thể viết như đoạn cuối cùng của phần *Thảo luận* vì nó mang tính chất chủ quan của tác giả, nhưng cũng có thể viết tách thành một mục riêng (thường là yêu cầu đối với Luận án tiến sĩ).

Kết luận là nơi tác giả trả lời các vấn đề hay câu hỏi nghiên cứu, đưa ra tuyên bố ủng hộ hay bác bỏ giả thuyết ban đầu. Kết luận phải rút ra được những ý tưởng chính, ý nghĩa của những phát hiện từ nghiên cứu và nêu được sự đóng góp cho khoa học, thực tiễn hay khả năng ứng dụng vào những lĩnh vực liên quan. Kết luận phải mang tính khái quát hóa, đảm bảo độ tin cậy. Kết luận phải nhất quán với kết quả thu được, trong đó cần nêu bật những phát hiện quan trọng và cả những hạn chế của nghiên cứu. Trong *Kết luận* của bài báo tác giả cũng nên phác thảo những ứng dụng khả dĩ hay những hàm ý của kết quả nghiên cứu.

Cuối cùng tác giả nên đề xuất ý tưởng cho các nghiên cứu trong tương lai để chính tác giả hay đồng nghiệp khác sẽ thực hiện nhằm phát triển những phát hiện mới hay để trả lời những câu hỏi nghiên cứu mới nảy sinh từ nghiên cứu hiện tại. Không đưa ra những “đề nghị” không xuất phát từ chính công trình nghiên cứu hiện tại.

k. Cảm tạ (Acknowledgements)

Phần này là nơi để tác giả ghi nhận và bày tỏ lời cảm ơn đối với những người và/hay cơ quan, tổ chức đã có đóng góp trong quá trình tiến hành nghiên cứu và/hay viết bài báo. Đó thường là những đóng góp về mặt tư vấn khoa học, hỗ trợ kỹ thuật trong quá trình thực hiện nghiên cứu hay tài trợ cho việc thực hiện công trình nghiên cứu. Không cảm ơn đồng tác giả đã có tên trong bài báo.

Ví dụ: “Công trình nghiên cứu này nằm trong khuôn khổ của Dự án A với sự tài trợ của Tổ chức B. Tác giả xin chân thành cảm ơn GS.TS. Nguyễn Văn C về những lời khuyên quý giá cho việc thiết kế thí nghiệm và những gợi ý thảo luận cho bài báo này. Chúng tôi cũng xin cảm ơn bà Trần Thị D đã giúp đỡ về kỹ thuật phân tích trong phòng thí nghiệm”.

l. Tài liệu tham khảo (References) và cách trích dẫn (Citation)

Trích dẫn tài liệu tham khảo là một yêu cầu bắt buộc đối với các bài viết khoa học, từ một bài tiểu luận, luận văn, luận án đến các báo cáo nghiên cứu khoa học, các bài viết đăng trên các tạp chí học thuật. Cách viết trích dẫn tài liệu tham khảo có ý nghĩa quan trọng trong việc đánh giá độ chuyên sâu, tính nghiêm túc và tính trung thực của tác giả. Các tài liệu tham khảo gồm sách, báo, tạp chí, ấn phẩm in và ấn phẩm điện tử của các cá nhân, tổ chức. Tác giả bài báo sử dụng số liệu, ý tưởng từ các công trình, tác phẩm của bất kỳ cá nhân hay tổ chức nào đều phải cung cấp đầy đủ thông tin nguồn gốc tài liệu được trích dẫn sao cho không vi phạm một điều cấm kỵ trong nghiên cứu khoa học – đạo văn.

Mỗi tạp chí có một quy định riêng về cách trích dẫn và trình bày tài liệu tham khảo. Do đó, trong thực tế, có rất nhiều cách trích dẫn và nhiều cách thức trình bày tài liệu tham khảo khác nhau. Tác giả phải nắm được chính xác và tuân thủ đúng quy định về cách trích dẫn và trình bày danh mục tài liệu tham khảo của tạp chí mà mình định gửi bài đăng. Tốt nhất là sử dụng một phần mềm như EndNote² để quản lý tài liệu tham khảo.

Nguyên tắc chung về trích dẫn và viết tài liệu tham khảo (TLTK) gồm:

- Tất cả những ý tưởng và dẫn liệu trong bài báo mà không phải của tác giả thì phải ghi rõ nguồn TLTK, trừ những kiến thức đã trở nên phổ biến (common

² EndNote là một phần mềm quản lý tài liệu tham khảo dùng để tổ chức và theo dõi các tài liệu tham khảo được sử dụng trong các bài viết. Sử dụng Endnote cho phép dễ dàng chèn các ghi chú trích dẫn hoặc chú thích vào tài liệu nghiên cứu đồng thời cho phép trao đổi thư mục. Sử dụng phần mềm Endnote giúp các nhà nghiên cứu tiết kiệm được thời gian khi lập danh mục tài liệu tham khảo cũng như tránh được những sơ suất không đáng có trong khi trình bày các bài báo cáo khoa học. Endnote là một phần mềm được đăng ký độc quyền, phát hành từ năm 1988 với dạng phần mềm đóng gói để có thể dễ dàng cài đặt trên máy tính đơn.

knowledge). Sử dụng ý tưởng, số liệu hay câu văn của người khác mà không ghi rõ nguồn gốc bị coi là đạo văn (plagiarism). Đó là một thứ “tội phạm” trong khoa học không thể chấp nhận được.

- Không trích dẫn những TLTK mà tác giả chưa đọc;
- Chỉ trích dẫn những tài liệu trực tiếp liên quan đến bài báo;
- Cách viết trích dẫn và trình bày danh mục TLTK tuân theo quy định của Tạp chí;
- Không trích dẫn TLTK trong phần trình bày Kết quả.

3.4. LUẬN ÁN TIẾN SĨ

Luận án tiến sĩ là một tài liệu khoa học trình bày công trình nghiên cứu gốc của nghiên cứu sinh, có đóng góp về mặt lý luận và thực tiễn trong lĩnh vực nghiên cứu hoặc giải pháp mới có giá trị trong việc phát triển, gia tăng tri thức khoa học của lĩnh vực nghiên cứu, giải quyết sáng tạo các vấn đề của ngành khoa học hay thực tiễn kinh tế - xã hội. Luận án là sản phẩm bắt buộc đối với mỗi NCS. Đó là văn bản gốc thể hiện năng lực tiến hành nghiên cứu độc lập của nghiên cứu sinh (dưới sự hướng dẫn của người hướng dẫn khoa học) và khả năng phân tích, truyền đạt kết quả có ý nghĩa của công trình đó.

3.4.1. Yêu cầu chung đối với luận án tiến sĩ

Luận án tiến sĩ phải đáp ứng được những yêu cầu chung của Quy chế đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo (Việt Nam) và Quy định đào tạo tiến sĩ của cơ sở đào tạo. Nói chung, một luận án tiến sĩ phải đáp ứng được các yêu cầu sau đây:

- Luận án phải phải *do chính nghiên cứu sinh thực hiện* và có những *đóng góp mới* về mặt học thuật, được trình bày bằng ngôn ngữ khoa học, vận dụng những lý luận cơ bản của ngành khoa học để phân tích, bình luận các luận điểm và kết quả đã đạt được trong các công trình nghiên cứu trước đây liên quan đến đề tài luận án, trên cơ sở đó đặt ra vấn đề mới, giả thuyết mới có ý nghĩa hoặc các giải pháp mới để giải quyết các vấn đề đặt ra của luận án và chứng minh được bằng những tư liệu mới.

- *Cấu trúc và thể thức* viết luận án tiến sĩ cũng tương tự như đối với một bài báo khoa học. Tuy nhiên, luận án có thể trình bày nhiều hơn một chủ đề và nhiều hơn một cách tiếp cận tới một vài chủ đề, có thể trình bày toàn bộ hay hầu như toàn bộ dữ liệu thu được từ đề tài nghiên cứu của NCS và phải theo quy định về thể thức của cơ sở đào tạo. Do vậy, luận án phải chi tiết và dài hơn một bài báo khoa học.

- Tác giả luận án phải có *cam đoan về công trình khoa học gốc* của mình và phải đảm bảo tuân thủ các quy định về *sở hữu trí tuệ*. Nếu luận án là công trình khoa học hoặc một phần công trình khoa học của một tập thể trong đó tác giả đóng góp phần chính thì phải xuất trình với cơ sở đào tạo các văn bản của các thành viên trong tập thể đó đồng ý cho phép nghiên cứu sinh sử dụng công trình này trong luận án để bảo vệ lấy Bằng Tiến sĩ. Việc sử dụng hoặc trích dẫn kết quả nghiên cứu của người khác, của đồng tác giả phải được dẫn nguồn đầy đủ và rõ ràng, trình bày theo đúng thể thức quy định chung của cơ sở đào tạo.

- Nội dung chủ yếu và các kết quả nghiên cứu của đề tài luận án phải được báo cáo tại các hội nghị khoa học chuyên ngành và được công bố trong các bài báo trên tạp chí khoa học chuyên ngành có phản biện độc lập. Số lượng *bài báo công bố* và danh mục các tạp chí chuyên ngành do cơ sở đào tạo quy định.

3.4.2. Cấu trúc của luận án

Hiện nay có hai dạng bố cục cơ bản cho một luận án tiến sĩ.

Dạng 1: Cấu trúc của luận án bao gồm các phần và chương

1. Phần phụ đầu luận án:

- Bìa và trang bìa phụ
- Lời cam đoan
- Lời cảm ơn
- Mục lục
- Danh mục chữ viết tắt
- Danh mục bảng
- Danh mục đồ thị, hình, sơ đồ, ảnh,...

2. Phần chính của luận án:

- Mở đầu
- Tổng quan tài liệu
- Phương pháp nghiên cứu
- Kết quả và thảo luận (có thể tách thành nhiều chương)
- Kết luận và kiến nghị

3. Danh mục các công trình công bố

4. Danh mục tài liệu tham khảo

5. Phần phụ lục (nếu có)

Dạng 2: Cấu trúc của luận án bao gồm các bài báo khoa học và các phần bổ sung

- Phần mở đầu: Giới thiệu chung về luận án
- Phần chính: Tập hợp các bài báo khoa học
- Thảo luận chung
- Phụ lục (nếu có)

Hiện nay hầu hết các cơ sở đào tạo ở Việt Nam đang có quy định cấu trúc luận án tiến sĩ theo dạng 1.

3.4.3. Nội dung và cách viết các phần chính của luận án

Nội dung và cách viết các phần chính của một luận án tiến sĩ về cơ bản cũng gần giống như các phần tương tự của một bài báo khoa học. Học viện Nông nghiệp Việt Nam có “Hướng dẫn cách viết và trình bày Luận án tiến sĩ”, mỗi NCS có trách nhiệm tìm hiểu chi tiết của hướng dẫn này để tuân thủ đúng quy định. Nội dung luận văn/luận án thông thường gồm nhiều phần và mỗi phần có thể cấu trúc như chương hay phần và các tiểu mục và tiểu mục phụ đi kèm. Sau đây là nội dung và cách viết các phần chính của luận án tiến sĩ.

a. Phần mở đầu

Phần mở đầu cung cấp thông tin ngắn gọn nhưng cần thiết để giúp cho người đọc hiểu ý tưởng nghiên cứu (vấn đề, câu hỏi, giả thuyết nghiên cứu), mục tiêu, phạm vi nghiên cứu, những đóng góp mới và ý nghĩa của đề tài luận án. Phần mở đầu thường chiếm khoảng 5% độ dài của luận án, trong đó tác giả phải trình bày các nội dung sau:

- Tính cấp thiết của đề tài luận án: Xác định vấn đề nghiên cứu, cụ thể là lập luận tại sao một vấn đề nào đó cần được quan tâm giải quyết và một số câu hỏi cụ thể cần được nghiên cứu trả lời. Do đó, phần này phải chứa đựng một số nội dung cốt lõi sẽ được đề cập tới ở chương Tổng quan tài liệu. Tác giả phải tổng hợp cô đọng, súc tích, có trích dẫn những phát hiện chủ yếu, đặc biệt những phát hiện mới nhất, để thể hiện được tổng quan ngắn gọn về những gì đã biết, những gì chưa biết, những gì còn tranh luận xung quanh vấn đề nghiên cứu để cuối cùng dẫn đến luận đề nghiên cứu (cần chứng minh điều gì?) nhằm lấp khoảng trống tri thức hiện tại hay tìm giải pháp cho vấn đề đang được quan tâm. Thông thường một đề tài luận án quan tâm đến một vấn đề “cấp thiết” của ngành/lĩnh vực chuyên môn, từ đó đặt ra một số câu hỏi nghiên cứu cho đề tài (nên 2-4) nhằm làm sáng tỏ một số khía cạnh cụ thể của vấn đề đó.

- Giả thuyết khoa học: Nêu giả thuyết khoa học (sự tiên đoán hay câu trả lời ước chừng) cho từng câu hỏi nghiên cứu và phác thảo ngắn gọn cách kiểm định giả thuyết đó trong phạm vi nghiên cứu của đề tài luận án.

- Mục tiêu nghiên cứu: Nêu những điểm cụ thể cần đạt được thông qua thực hiện đề tài luận án nhằm kiểm định từng giả thuyết hay giải quyết vấn đề/câu hỏi nghiên cứu. Do đó các mục tiêu nghiên cứu cụ thể phải logic với tính cấp thiết và (các) câu hỏi nghiên cứu/giả thuyết khoa học đã nêu.

- Những đóng góp mới: Luận án bắt buộc phải có những *đóng góp mới* về mặt học thuật, do đó tác giả phải nêu bật được đâu là tính mới của đề tài luận án. Chú ý, đây không phải là những gì lần đầu tiên được thực hiện (new doings) ở đâu đó như nhiều luận án vẫn nêu mà phải là những phát hiện mới (new findings).

- Ý nghĩa khoa học và thực tiễn: Nêu rõ những phát hiện mới từ đề tài luận án có góp phần gì vào việc lấp khoảng trống tri thức hiện tại (ý nghĩa khoa học) hay tìm giải pháp hữu ích cho vấn đề đang được quan tâm, đặc biệt là những vấn đề nóng nảy sinh trong thực tiễn (ý nghĩa thực tiễn).

b. Tổng quan tài liệu

Khác với bài báo khoa học trong luận án tiến sĩ cũng như luận văn thạc sĩ có một chương riêng về tổng quan tài liệu. Chương này nhằm phân tích, đánh giá có biện luận các công trình nghiên cứu đã được công bố (trong và ngoài nước) có liên quan mật thiết đến đề tài luận án; chỉ ra được những gì đã biết, đã được giải quyết cũng như những gì chưa biết, còn tồn tại mà luận án tập trung giải quyết. Vì đề tài luận án là một nghiên cứu gốc nhằm đóng góp vào tri thức khoa học hiện tại hay tri thức nghề nghiệp nên tổng quan tài liệu phải thiết lập được tính mới cho luận án. Nguyên tắc chung là “biết thì nói là biết, không biết thì nói là không biết, vậy mới thật là biết” (Khổng Tử, 551-479 TCN), nghĩa là trong chương này nghiên cứu sinh phải thể hiện mình “thật là biết” về những gì đã biết và cả những gì chưa biết xung quanh vấn đề quan tâm nhằm biện minh cho sự cần thiết và định hướng cho đề tài nghiên cứu của mình.

Tổng quan tài liệu chiếm khoảng 25-30% của luận án, trong đó thông qua tổng hợp tài liệu và tư duy biện luận, nghiên cứu sinh cần:

- Xác định những khoảng trống trong tri thức hiện tại về vấn đề nghiên cứu;
- Chỉ rõ những nhận thức về những phát triển quan trọng trong lĩnh vực đó;
- Nêu rõ những quan điểm khác nhau và những vấn đề còn tranh luận;
- Đánh giá có biện luận những điểm mạnh và điểm yếu của các nghiên cứu trước của các tác giả khác và cả của chính mình (nếu có);
- Chứng minh nghiên cứu của mình là sự tiếp nối/dựa trên những công trình trước đây và đề ra điểm xuất phát của vấn đề nghiên cứu của chính mình;
- Chứng minh tầm quan trọng và sự phù hợp của đề tài luận án;
- Định hướng giải quyết vấn đề nghiên cứu cho đề tài luận án.

Nghiên cứu sinh cần tránh viết tổng quan tài liệu theo kiểu liệt kê rời rạc các công trình nghiên cứu trước đây mà không sắp xếp theo một thiết kế tổ chức ý tưởng nào cả hay không có đánh giá, biện luận của chính mình. Mặt khác, những kiến thức chung và phổ biến như trong giáo trình và sách phổ biến kiến thức chuyên môn cũng không đưa vào phần này vì yêu cầu là phải cập nhật những kiến thức mới nhất trong lĩnh vực nghiên cứu hẹp để đảm bảo rằng đề tài luận án không lặp lại mà chỉ là sự tiếp nối nhằm bổ sung thêm kiến thức mới.

Để viết tổng quan tài liệu, nghiên cứu sinh phải sưu tầm tài liệu, tổng hợp, đánh giá có biện luận các công trình đã được tiến hành, nhất là những công trình nổi bật trong thời gian gần đây trong phạm vi nghiên cứu của đề tài, từ đó rút ra những điểm chính và dự đoán xu hướng của các nghiên cứu trong tương lai. Các nội dung trong tổng quan tài liệu phải được sắp xếp một cách logic xung quanh một luận đề chính nhằm cung cấp cho người đọc những kiến thức cần thiết liên quan đến câu hỏi nghiên cứu và mục tiêu của đề tài luận án. Cần phải nhóm lại thành các nhóm ý tưởng của các tác giả khác nhau, trên cơ sở đó hình thành những quan điểm/đánh giá của chính mình. Thông

thường thì tổng quan tài liệu được thiết kế theo dạng hình phễu: bắt đầu ở mức tổng quát, thu hẹp dần để dẫn dắt đến từng câu hỏi nghiên cứu.

Về mặt cấu trúc viết, một tổng quan tài liệu là một chỉnh thể thống nhất và có hình thức giống như một bài báo khoa học tổng quan (review paper). Mặc dù không có một quy định thống nhất nào, nhưng thông thường thì chương tổng quan tài liệu gồm có phần mở đầu/giới thiệu, phần thân và phần kết luận; tất cả gắn kết với nhau một cách logic xoay quanh một trục luận điểm chính hướng tới câu hỏi nghiên cứu. Phần mở đầu viết ngắn gọn nhằm giới thiệu vấn đề nghiên cứu và dàn ý của chương tổng quan tài liệu. Phần thân được chia thành nhiều tiểu mục nhằm tổng hợp và đánh giá hiện trạng tri thức về các khía cạnh khác nhau xung quanh vấn đề nghiên cứu (liên quan đến các mục tiêu và nội dung nghiên cứu của đề tài). Phần kết luận tóm tắt lại những bằng chứng đã trình bày, ý nghĩa của chúng và dẫn ra những khoảng trống kiến thức hiện tại (những gì chưa biết), từ đó định hướng cho việc giải quyết vấn đề nghiên cứu thông qua đề tài luận án như là một sự tiếp nối các công trình nghiên cứu trước đó nhằm bổ sung kiến thức mới.

c. Vật liệu và phương pháp nghiên cứu

Chương/phần này nhằm mô tả cách tiến hành thí nghiệm, thực nghiệm, điều tra,... để thu thập thông tin, số liệu trả lời câu hỏi nghiên cứu đặt ra. Về căn bản, phần này cần mô tả theo trình tự logic: nội dung nghiên cứu, cách thiết kế và tiến hành nghiên cứu, những vật liệu và phương pháp nào đã được sử dụng, cách thu thập, xử lý và phân tích số liệu. Nói một cách khác, chương/phần này nhằm trả lời được những câu hỏi: Làm những gì (what)? Làm như thế nào (how)? Tại sao lại sử dụng phương pháp đó (why)? Tiến hành khi nào (when)? Và ở đâu (where)? Để đạt được các mục tiêu cụ thể của đề tài.

Do đó, chương/phần này thường chiếm tối đa 15% độ dài của luận án và có các tiểu mục sau:

- Đối tượng/Vật liệu nghiên cứu. Đối tượng nghiên cứu là sự vật, hiện tượng được lựa chọn để xem xét trong đề tài nghiên cứu. Cách viết mục này tương tự như đối với bài báo khoa học (xem mục 3.3.3).

- Phạm vi nghiên cứu. Phạm vi nghiên cứu là giới hạn nghiên cứu của đề tài, bao gồm giới hạn về đối tượng, thời gian và không gian/địa điểm nghiên cứu.

- Nội dung nghiên cứu. Trả lời câu hỏi đã làm những gì để đạt được mục tiêu của đề tài luận án này? Do đó, nội dung nghiên cứu phải phù hợp với mục tiêu của đề tài: có mục tiêu nghiên cứu thì phải có nội dung nghiên cứu để đạt được, và ngược lại. Do đó các nội dung nghiên cứu phải được sắp xếp tương ứng với từng mục tiêu cụ thể của đề tài, có bao nhiêu mục tiêu thì có bấy nhiêu nhóm nội dung nghiên cứu.

- Phương pháp nghiên cứu. Cách viết mục này tương tự như đối với bài báo khoa học (xem mục 3.3.3) nhằm làm rõ: (1) Thiết kế thí nghiệm, (2) Quy trình thực hiện, (3) Phương pháp thu thập số liệu (trong quá trình thí nghiệm và cách tính toán các chỉ tiêu nghiên cứu) và (4) Phương pháp xử lý thống kê số liệu. Việc mô tả phương pháp phải

đầy đủ, chi tiết, rõ ràng và cụ thể để đồng nghiệp có thể tham khảo và lặp lại được nghiên cứu một cách chính xác.

d. Kết quả và thảo luận

Kết quả và thảo luận có thể trình bày trong cùng chương hay cùng trong mỗi mục/tiểu mục, trình bày kết quả trước và thảo luận sau đó. Tuy nhiên, Học viện Nông nghiệp Việt Nam khuyến khích (không bắt buộc) nghiên cứu sinh viết tách kết quả (mang tính khách quan) và thảo luận (mang tính chủ quan) thành 2 phần hay chương riêng. Phần kết quả và thảo luận thường chiếm tối thiểu 50% độ dài của luận án. Nguyên tắc chung để viết kết quả và thảo luận cũng tương tự như đối với bài báo khoa học (xem Mục 3.3.3).

Kết quả phải được trình bày được chính xác và nêu bật được những phát hiện chính của đề tài và phải được tổ chức theo trình tự thông tin phù hợp với câu hỏi/mục tiêu nghiên cứu đề ra. Do đó, cần trình bày kết quả *theo các tiểu mục* tương ứng với các nội dung nghiên cứu nhằm trả lời từng câu hỏi và mục tiêu nghiên cứu. Mặt khác, một điều cần đặc biệt lưu ý là các kết quả trình bày trong luận án phải nhất quán với những kết quả đã được công bố trước đó trong các bài báo như là một điều kiện để bảo vệ luận án.

Thảo luận là phần được coi là khó nhất đối với nghiên cứu sinh vì nó đòi hỏi tác giả phải am hiểu về lĩnh vực nghiên cứu để biết cách giải thích kết quả, trả lời câu hỏi nghiên cứu, biện minh cho phương pháp và đánh giá có biện luận nghiên cứu của mình. Hơn thế, tác giả phải giải thích và khái quát hoá được kết quả nghiên cứu, đưa ra những luận điểm, nhận thức mới và chỉ ra ý nghĩa của các phát hiện mới. Việc đưa ra *luận điểm* phải dựa trên *luận cứ* cụ thể, đáng tin cậy với các *luận chứng* hợp lý³. Ngoài ra, tác giả cần phân tích, nhận biết được cả những hạn chế và tồn tại của đề tài luận án.

e. Kết luận và kiến nghị

Kết luận phải rút ra được những câu trả lời cho các câu hỏi nghiên cứu, chấp nhận hay bác bỏ các giả thuyết của đề tài và nêu được sự đóng góp cho tri thức khoa học, thực tiễn hay khả năng ứng dụng vào những lĩnh vực liên quan. Kết luận phải mang tính khái quát hóa, đảm bảo độ tin cậy và có giá trị trên cơ sở suy luận quy nạp từ chính kết quả nghiên cứu của đề tài luận án (với bằng chứng cụ thể). Tránh viết kết luận theo kiểu tóm tắt lại các kết quả nghiên cứu (nêu những con số) mà phải làm rõ những kết quả đó nói lên điều gì theo mục tiêu đặt ra của đề tài. Phần kết luận cũng có thể gồm những hạn chế của nghiên cứu.

Kiến nghị phải đưa ra những đề xuất cụ thể cho nghiên cứu tiếp theo để trả lời cho những câu hỏi nghiên cứu mới nảy sinh từ đề tài luận án. Tác giả cũng có thể đề nghị sử dụng những kết quả nghiên cứu mới từ đề tài luận án. Các kiến nghị phải nêu cụ thể như là sản phẩm có ý nghĩa của đề tài luận án, chứ không nêu chung chung dạng như “Đề

³ Luận điểm là ý kiến của người viết về vấn đề/câu hỏi được đặt ra. *Luận cứ* là những bằng chứng, lý lẽ được dùng để thuyết minh cho luận điểm. *Luận chứng* là cách thức phối hợp, tổ chức các bằng chứng và lý lẽ để thuyết minh cho luận điểm.

ngộ tiếp tục nghiên cứu vấn đề này để có kết luận chính xác hơn” (người đọc sẽ hiểu là đề tài luận án thực hiện chưa được tốt) hay “Đề nghị cho áp dụng các kết quả nghiên cứu của đề tài này vào sản xuất” (không biết áp dụng cái gì vào cái gì?).

3.5. VĂN PHONG VÀ ĐẠO VĂN TRONG CÔNG BỐ KHOA HỌC

3.5.1. Văn phong khoa học

Thí nghiệm tôi không thể che giấu được nhờ viết tốt, nhưng viết tôi sẽ làm lu mờ thí nghiệm tốt. Nội dung là quan trọng, nhưng trình bày là cốt tử. Một bài báo khoa học tốt phải làm cho người đọc nhận ngay ra được ý nghĩa khoa học của nó. Mục đích của viết tài liệu khoa học là để chuyển tải chính xác thông tin khoa học, bởi vậy nó phải được viết một cách đơn giản, sử dụng ngôn ngữ phù hợp với số lượng từ ít nhất. Nói một cách khác, “khoa học tốt phải được chuyển tải tốt”. Bài viết phải chuyển tải được các thông điệp khoa học hữu ích một cách hấp dẫn, truyền đạt được suy nghĩ của tác giả theo một trình tự logic sao cho người đọc rút ra được những kết luận giống như ý của tác giả. Do vậy, một bài báo khoa học tốt thì ngoài số liệu (nguyên liệu) tốt còn phải được cấu trúc theo một thể thức phù hợp và phải được viết theo văn phong khoa học.

Văn phong khoa học khác với văn phong của tiểu thuyết, thơ ca (văn học) hay văn báo chí (newspapers). Trong văn học tác giả có thể viết dài dòng, bóng bẩy, tình tế, huyền ảo, bóng gió, lơ lửng, hững hờ và ngụ ý để chuyển tải một thông điệp tiềm thức và cảm tính, do đó người đọc có thể suy ngẫm, tưởng tượng và hiểu theo cách riêng của mình. Nhà văn phản ánh hiện thực thông qua “lăng kính” riêng của mình. Các nhà báo nhiều khi chủ ý chỉ viết một phần của sự thật. Văn khoa học phải thể hiện sự minh bạch, thực tế, chính xác, dứt khoát và gọn gàng theo những nguyên tắc chặt chẽ. Mỗi bài báo khoa học hay luận án là một tài liệu hoàn chỉnh mà tác giả phải viết sao cho người đọc có thể lĩnh hội đúng thông điệp mình muốn chuyển tải. Do vậy, các thông tin phải được sắp xếp một cách logic và khúc chiết theo trình tự của một câu chuyện được kể từ đầu đến cuối tuân thủ theo một hình thức được quy định rất chặt chẽ.

Văn khoa học không cho phép dùng những câu từ phức tạp, khó hiểu hay có thể hiểu theo nhiều nghĩa khác nhau. Các từ ngữ trong bài báo phải chính xác, đơn giản, trực tiếp và dễ hiểu. Tác giả phải sử dụng từ, nhất là các thuật ngữ khoa học, một cách phù hợp và *nhất quán* trong toàn bài. Muốn kết quả nghiên cứu của mình được nhiều người hiểu và chia sẻ thì tác giả cũng nên tránh dùng các biệt ngữ, từ hoa mỹ, từ tối nghĩa hay mơ hồ.

Câu văn khoa học nên cố gắng viết ngắn gọn. Loại câu thường dùng là câu trần thuật có đầy đủ chủ ngữ và vị ngữ. Tránh dùng những câu văn không có hay không rõ chủ ngữ. Ví dụ, nhiều tác giả hay bắt đầu câu văn dạng “*Theo bảng X cho thấy ...*” hay “*Theo Preston (2013) cho thấy ...*”. Những câu như vậy về văn phạm là không có chủ ngữ vì nó có giới từ “*theo*” đi trước danh từ, mặc dù người đọc có thể hiểu được rằng *bảng X* hay *Preston* là chủ thể của “*cho thấy*”. Có thể sửa lại hai câu vừa rồi bằng cách bỏ đi chữ “*theo*” ở đầu câu hay bỏ “*cho thấy*” và tiếp tục câu (...).

Về thể văn và đại từ nhân xưng, trước đây nhiều tạp chí khoa học có chủ trương dùng thể văn bị động (passive voice) và tránh dùng đại từ nhân xưng ngôi thứ nhất (tôi, chúng tôi) làm chủ ngữ nhằm thể hiện tính khách quan trong khoa học. Ví dụ, viết “*Thí nghiệm được tiến hành...*”/“*The experiment was conducted ...*”, chứ không viết “*Chúng tôi tiến hành thí nghiệm...*”/“*We conducted the experiment ...*”. Tuy nhiên, gần đây một số tạp chí lớn lại khuyến khích tác giả nên chọn thể văn chủ động (active voice) vì thể văn này thể hiện được tính sống động hơn cho bài báo và tránh được những câu văn rờn rờ. Hơn nữa, nếu bài báo viết bằng tiếng Việt thì cách viết câu bị động (kèm theo từ “bị” hay “được” trước động từ) là không phổ biến nhiều khi rất khó viết. Mặt khác, tác giả cũng phải “chịu trách nhiệm cá nhân” trong khoa học nên việc dùng đại từ nhân xưng ngôi thứ nhất là có lý, nhất là khi thể hiện quan điểm của tác giả (chưa chắc người khác đã tán thành). Ví dụ, nên viết “*Chúng tôi cho rằng...*” để thể hiện sự nhìn nhận của các tác giả về một kết quả nào đó. Do vậy, trong những phần mang nhiều tính chủ quan như phần *Thảo luận*, kể cả *Kết luận*, trong nhiều tạp chí hiện nay người ta hay viết câu chủ động và dùng đại từ nhân xưng làm chủ ngữ. Ví dụ, thay vì viết “*Kết luận được rút ra là ...*”/“*It was concluded that...*” có thể viết “*Chúng tôi kết luận rằng ...*”/“*We conclude that ...*”.

Thì của động từ cũng cần được chú ý, nhất là khi viết bài báo hay tóm tắt bài báo bằng tiếng Anh. Trong văn khoa học chỉ có 2 thì của động từ được dùng chủ yếu là thì hiện tại đơn giản (simple present tense) và thì quá khứ đơn giản (đã + động từ/simple past tense). Thì hiện tại được dùng cho những kiến thức phổ quát hay mô tả bảng biểu trong bài báo. Thì quá khứ được dùng để nói về những kết quả nghiên cứu cụ thể trước đây (trích dẫn từ tài liệu tham khảo), mô tả kết quả của nghiên cứu gốc trong bài báo. Do vậy, trong phần *Phương pháp nghiên cứu* và phần *Kết quả* thì nên dùng động từ ở thì quá khứ để báo cáo những việc đã làm và kết quả đã thu được; trong phần *Đặt vấn đề*, phần *Tổng quan tài liệu* và phần *Thảo luận/Kết luận* có thể dùng cả thì quá khứ và thì hiện tại tùy theo văn cảnh.

Đoạn văn (paragraph) là đơn vị viết cốt yếu để diễn đạt ý nghĩ. Đoạn văn khoa học gồm một số câu văn hợp lại để nói lên được một điểm thông tin/chủ đề chính. Một đoạn văn khoa học thường chỉ gồm 5 đến 7 câu, trong đó có một câu chủ đề (topic sentence) và một số câu bổ trợ (supportive sentences). Đôi lúc (không phải tất cả) cuối đoạn văn có một câu kết tóm tắt lại các thông tin đã nói trong đoạn văn (nhất là khi đoạn văn dài) và thường được viết như là câu chủ đề đảo ngược (không phải viết giống hệt).

Câu văn chủ đề nhằm thông báo với người đọc một ý tưởng chủ đạo (controlling idea) của đoạn văn. Nó thường được viết đầu đoạn văn và là câu khái quát nhất trong đoạn văn thể hiện ý chủ đạo/bao trùm cho nội dung sẽ được thảo luận sau đó trong đoạn văn. Không đi vào chi tiết trong câu này. Ví dụ: Lợn Móng Cái là một giống lợn quý nhờ một số ưu điểm nổi bật (chưa nêu cụ thể đó là những gì mà các câu bổ trợ tiếp theo sẽ diễn giải cụ thể).

Những câu văn bổ trợ có chức năng giải thích thêm hay chi tiết hoá cho điểm chính/ý tưởng chủ đạo. Thường có nhiều hơn hai ý/câu bổ trợ, một ý bổ trợ có thể thể hiện bằng nhiều câu. Nếu có thể thì cung cấp đủ chi tiết cho mỗi ý bổ trợ để người đọc có thể hiểu chính xác người viết muốn nói gì. Những câu bổ trợ này thường nhằm mục đích:

- Cung cấp ví dụ để minh họa;
- Cung cấp thêm chi tiết để diễn giải điểm chính;
- Nhấn mạnh đến ý nghĩa của điểm chính;

Để người đọc dễ theo dõi, đoạn văn phải được viết trôi chảy. Muốn vậy, các câu văn phải gắn kết chặt chẽ với nhau, có thể thông qua các liên từ hay phối hợp “ăn ý” với nhau. Mỗi câu văn đi sau là sự mở rộng hay nối ý cho câu văn trước. Tương tự như giữa các câu trong một đoạn văn, các đoạn văn cũng phải được nối kết với nhau một cách logic, phối hợp với nhau để thể hiện được thông điệp chính của toàn bài. Không được viết một đoạn văn gồm quá nhiều câu về những ý tưởng rời rạc, không ăn khớp với nhau. Có những nghiên cứu sinh viết những đoạn văn trong Luận án dài hơn cả trang giấy A4, không cho người đọc có thời gian “giải lao”. Trái lại, nhiều người cứ viết hết mỗi câu lại xuống dòng, buộc người đọc phải “nhặt nhạnh” các ý tứ rải rác mà không biết tác giả thực sự muốn nói gì.

Bảng 3.6. Lí do bài báo khoa học bị từ chối công bố

Lý do chính	Phần trăm (%)
Phương pháp và phương pháp luận	74,3
Tầm ảnh hưởng thấp hay không có	60,3
Văn phong	58,4
Tổng quan tài liệu	50,9
Phân tích dữ liệu	42,1
Cấu trúc bài báo	34,6
Chất lượng nghiên cứu và tính nghiêm túc	30,0
Lấy mẫu	29,2
Kết luận	27,6
Thảo luận	25,2
Tài liệu tham khảo	23,6

(Nguyễn Văn Tuấn, 2012)

Tóm lại, muốn một tài liệu công bố khoa học tốt, ngoài nội dung tốt còn phải đáp ứng được một số tiêu chuẩn về cách viết. Theo Giáo sư Nguyễn Văn Tuấn (2012) ở Đại học New South Wales (Australia), tác giả của trên 200 bài báo khoa học y khoa và có kinh nghiệm lâu năm làm thành viên ban biên tập của 7 tạp chí y khoa quốc tế, văn phong là một trong 3 nguyên nhân thường gặp nhất để phản biện từ chối công bố bài báo (Bảng 3.6). Những tiêu chuẩn về cách viết một bài báo khoa học tốt bao gồm: cách

phát biểu mục đích, tuân thủ quy định của tạp chí, dùng từ ngữ chính xác và nhất quán, cách trình bày logic và khúc chiết, dữ liệu được trình bày qua những bảng biểu có chất lượng cao (Bảng 3.7).

Bảng 3.7. Tiêu chuẩn cho một bài báo khoa học tốt

Tiêu chuẩn	Mô tả
Mục đích cụ thể	Phát biểu rõ ràng, cụ thể để người đọc biết công trình nghiên cứu muốn giải quyết vấn đề gì
Tuân thủ	Hình thức bài báo phải tuân thủ theo những quy định của Tạp chí
Từ ngữ chính xác	Từ ngữ dùng trong bài phải chính xác, đơn giản; không có sai sót về văn phạm
Nhất quán	Các thuật ngữ phải được dùng một cách thích hợp và nhất quán trước sau
Logic và khúc chiết	Nên viết theo thể văn kể chuyện theo trình tự: đặt vấn đề, phương pháp, kết quả, thảo luận. Bài báo có một thông điệp chính để người đọc có thể lĩnh hội
Dữ liệu tốt	Được trình bày qua những bảng biểu có chất lượng cao

(Nguyễn Văn Tuấn, 2012)

3.5.2. Đạo văn trong công bố khoa học

Đạo văn (plagiarism) là công bố ngôn ngữ, suy nghĩ, ý tưởng, hay cách diễn đạt của người khác như là những gì do mình tự tạo ra. Sao chép bất cứ thứ gì mà không ghi rõ tác giả của nó đều bị coi là đạo văn. Chẳng hạn như việc sử dụng ý tưởng hay câu văn của người khác mà không ghi rõ nguồn gốc, đặc biệt là việc trình bày những ý tưởng và từ ngữ của người khác trước các diễn đàn khoa học và công cộng như là ý tưởng và từ ngữ của chính mình. Xuất bản lại cùng dữ liệu đã công bố của chính mình cũng bị coi là đạo văn.

Theo Merriam-Webster Dictionary (<http://www.merriam-webster.com>), đạo văn nghĩa là:

- Ăn cắp và hình thành những ý tưởng hay ngôn từ mới khởi nguồn từ ý tưởng của ai đó;
- Sử dụng sản phẩm của một ai đó mà không công bố nguồn;
- Giới thiệu một ý tưởng hay sản phẩm mới được chuyển hóa từ một nguồn đã có từ trước.

Thậm chí mặc dù người viết đã dẫn nguồn nhưng vẫn bị coi là đạo văn trong những trường hợp sau:

- Dẫn tên tác giả nhưng không điền thông tin trích dẫn cụ thể để dẫn chứng về đoạn dẫn nguồn tham khảo như năm xuất bản, trang, chương mục...
- Cung cấp thông tin sai sự thật liên quan đến các nguồn tham khảo, khiến độc giả không thể tìm thấy được nguồn chính xác
- Có dẫn nguồn nhưng không có dấu trích dẫn khi đoạn đó được sao chép nguyên văn

- Có chỉ dẫn nguồn ở một vài nội dung tham khảo nhưng không trích dẫn tiếp khi sử dụng các nội dung khác của cùng một nguồn

Đạo văn được xem là hành vi thiếu trung thực về mặt học thuật. Trong môi trường học thuật và công việc, đạo văn là một hành vi vi phạm đạo đức rất nghiêm trọng; một số trường hợp đạo văn có thể cấu thành hành vi vi phạm bản quyền. Đạo văn được coi là lỗi nghiêm trọng, có thể chôn vùi sự nghiệp của bất cứ ai, không có ngoại lệ hay bất cứ biện hộ nào được chấp nhận. Người nào đạo văn sẽ bị phạt, bị đình chỉ và thậm chí bị đuổi học hay đuổi việc. Do đó, một khi đã có ý định sử dụng trích dẫn từ thành quả sáng tạo và lao động của người khác, tác giả phải ghi rõ nguồn và tên trích dẫn từng đoạn/từng ý một. Trong một số trường hợp cần phải xin phép.

Có rất nhiều công cụ phát hiện đạo văn online, chẳng hạn như phần mềm CopyCatch Gold, các trang web tìm kiếm như Google hay AltaVista... Hiện nay, Turnitin là trang phát hiện đạo văn được sử dụng rộng rãi tại các trường đại học trên thế giới. Ở nhiều trường đại học, giáo viên luôn yêu cầu sinh viên nộp tiểu luận, luận văn hay luận án của mình qua một chương trình chống đạo văn nào đó. Chỉ có những sinh viên đã qua được “kiểm duyệt đạo văn” đó mới được phép gửi bài đến giáo viên qua hình thức điện tử hay bản in.

CÂU HỎI ÔN TẬP

1. Nêu các phương pháp trình bày kết quả nghiên cứu?
2. Thế nào là bảng số liệu? Cách trình bày?
3. Biểu đồ là gì? Cách chọn loại biểu đồ để trình bày kết quả nghiên cứu?
4. Nguyên tắc dùng văn viết mô tả số liệu?
5. Phương pháp chuẩn bị báo cáo khoa học trình bày trong hội nghị khoa học?
6. Nêu cách thuyết trình báo cáo khoa học trình bày trong hội nghị khoa học?
7. Yêu cầu và cách thức trả lời chất vấn sau thuyết trình báo cáo khoa học?
8. Khái niệm và phân loại bài báo khoa học?
9. Cấu trúc của bài báo khoa học gốc?
10. Cần chuẩn bị những gì trước khi viết bài báo khoa học?
11. Cách viết từng phần của bài báo khoa học?
12. Luận án tiến sĩ là gì? Sự giống và khác nhau giữa luận án tiến sĩ và bài báo khoa học?
13. Yêu cầu chung đối với luận án tiến sĩ?
14. Cấu trúc của luận án tiến sĩ?
15. Nội dung và cách viết các phần chính của luận án?
16. Nêu nhưng yêu cầu cơ bản của văn phong khoa học.
17. Thế nào là đạo văn trong công bố khoa học? Tính chất nghiêm trọng của đạo văn?

Chương 4

THỰC HÀNH PHÂN TÍCH VÀ CÔNG BỐ KẾT QUẢ THÍ NGHIỆM

Theo Quy chế đào tạo tiến sĩ phần quan trọng nhất trong chương trình đào tạo mà mọi nghiên cứu sinh đều phải thực hiện là đề tài nghiên cứu khoa học. Nghiên cứu sinh không những phải biết thực hiện thí nghiệm, mà phải có kỹ năng phân tích kết quả và công bố kết quả nghiên cứu. Do vậy, mục đích của Chương này là giúp nghiên cứu sinh “học qua hành” để củng cố các kỹ năng quan trọng đó sau khi đã có các kiến thức học từ các chương trước.

4.1. NỘI DUNG THỰC HÀNH

Mỗi nghiên cứu sinh được cho sẵn một bộ số liệu của một thí nghiệm chăn nuôi để thực hành xử lý và phân tích thống kê. Nghiên cứu sinh cũng có thể thực hành trên bộ số liệu của chính mình nhưng phải được giảng viên đồng ý. Sau đó, mỗi nghiên cứu sinh viết bản thảo cho một bài báo khoa học để công bố kết quả và chuẩn bị một báo cáo PowerPoint trình bày dưới dạng seminar khoa học. Sản phẩm mỗi nghiên cứu sinh cần nộp cho giảng viên phụ trách để đánh giá gồm:

1. Kết quả chạy phần mềm máy tính phân tích thống kê bộ dữ liệu thí nghiệm (ghi rõ các câu lệnh và bản copy kết quả phân tích của phần mềm máy tính);
2. Bản thảo bài báo công bố kết quả gửi đăng trên Tạp chí của Học viện Nông nghiệp Việt Nam;
3. Bài báo cáo bằng PowerPoint và trình bày seminar.

4.2. BÀI THỰC HÀNH

4.2.1. Thí nghiệm

Một thí nghiệm được tiến hành tại Khoa Chăn nuôi, Học viện Nông nghiệp Việt Nam, nhằm đánh giá khả năng thay thế thức ăn tinh hỗn hợp hoàn chỉnh bằng thức ăn xanh giàu protein trong khẩu phần nuôi thỏ thịt New Zealand. Tổng số 35 thỏ đực New Zealand 1,5 tháng tuổi được phân ngẫu nhiên vào 5 lô, mỗi lô 7 con, để cho ăn 5 khẩu phần ăn thí nghiệm được thiết kế với các mức thay thế thức ăn viên hỗn hợp thương phẩm nuôi thỏ bằng một loại thức ăn xanh giàu protein (chè đại, rau muống, rau lang hoặc một loại lá giàu protein khác sẽ được xác định vào thời điểm thí nghiệm). Khẩu phần cho các lô cụ thể như sau:

- Lô 1: Thức ăn viên hỗn hợp chuyên dùng cho thỏ (đối chứng)
- Lô 2: Thay thế 25% thức ăn viên hỗn hợp bằng thức ăn xanh giàu protein
- Lô 3: Thay thế 50% thức ăn viên hỗn hợp bằng thức ăn xanh giàu protein
- Lô 4: Thay thế 75% thức ăn viên hỗn hợp bằng thức ăn xanh giàu protein

Lô 5: Ăn tự do thức ăn xanh giàu protein (không có thức ăn tinh)

Kết quả theo dõi khối lượng của thỏ theo thời gian thí nghiệm được trình bày ở bảng 4.1.

Bảng 4.1. Khối lượng tích lũy của thỏ thí nghiệm (g)

Thỏ	Lô	Ngày thí nghiệm									
		0	7	14	21	28	35	42	49	56	63
1	1	1600	1820	2110	2360	2480	2580	2600	2690	2820	2770
2	1	1480	1620	1920	2280	2250	2500	2500	2620	2740	2820
3	1	1680	1920	2230	2410	2440	2530	2500	2660	2680	2700
4	1	1280	1270	1510	1550	1820	1870	1900	2050	2080	2458
5	1	1700	1880	2110	2360	2380	2460	2460	2500	2540	2600
6	1	1540	1680	1900	2280	2420	2540	2560	2750	2800	2830
7	1	1740	1930	2100	2280	2360	2420	2380	2480	2564	2678
8	2	1620	1670	1930	2100	2200	2360	2340	2500	2500	2600
9	2	1500	1650	2030	2330	2530	2680	2740	2780	2840	2860
10	2	1340	1500	1810	1980	2060	2140	2200	2300	2300	2300
11	2	1500	1640	1860	2000	2100	2180	2200	2320	2340	2420
12	2	1820	1900	2250	2480	2700	2750	2880	2940	2976	3001
13	2	1360	1420	1650	1800	2020	2200	2240	2430	2400	2460
14	2	1820	1920	2140	2240	2370	2420	2410	2450	2520	2470
15	3	1400	1500	1790	1870	1990	2070	2120	2300	2380	2440
16	3	1500	1600	1950	2000	2160	2200	2300	2320	2400	2450
17	3	1400	1470	1710	1880	2030	2120	2140	2170	2200	2230
18	3	1560	1610	1930	2050	2180	2290	2380	2460	2540	2580
19	3	1720	1800	2020	2200	2380	2380	2500	2660	2660	2640
20	3	1500	1640	1860	1950	2080	2120	2200	2280	2300	2210
21	3	1680	1740	2050	2100	2240	2240	2380	2400	2470	2500
22	4	1560	1600	1870	1940	2170	2230	2300	2340	2300	2260
23	4	1900	1820	2170	2230	2430	2560	2580	2603	2783	2876
24	4	1480	1430	1630	1700	1850	1870	1900	2060	2100	2000
25	4	1320	1340	1620	1680	1920	1960	2030	2150	2200	2100
26	4	1380	1450	1610	1770	1860	1900	2100	2240	2300	2440
27	4	1660	1700	1890	1980	2200	2140	2200	2300	2270	2310
28	4	1740	1680	1910	2000	2250	2380	2460	2620	2500	2500
29	5	1820	1700	1940	2020	2270	2320	2490	2380	2550	2530
30	5	1440	1500	1670	1830	2120	2100	2340	2300	2340	2420
31	5	1440	1450	1600	1750	2110	2000	1860	2060	2103	2152
32	5	1620	1510	1800	1850	1930	1880	2020	2060	2160	2231
33	5	1340	1310	1490	1540	1640	1680	1760	1920	1900	1956
34	5	1560	1450	1640	1510	1900	1900	2070	2100	2100	2200
35	5	1660	1560	1690	1800	1860	1940	1960	2000	2134	2253

4.2.2. Yêu cầu thực hành

Căn cứ vào mô tả thí nghiệm và kết quả ở bảng 4.1 về khối lượng tích lũy của thỏ qua các tuần thí nghiệm, mỗi nghiên cứu sinh tự thực hiện và nộp kết quả thực hành theo một số yêu cầu sau đây:

1. Xác định mô hình thiết kế của thí nghiệm đã cho;
2. Xử lý số liệu thô trên phần mềm Excel và tính các chỉ tiêu sau cho từng cá thể:
 - Tăng khối lượng bình quân = hệ số hồi quy (slope) của khối lượng theo thời gian thí nghiệm
 - Tăng khối lượng cả kỳ = KL cuối kỳ - KL đầu kỳ
3. Vẽ đồ thị và viết phương trình hồi quy tốt nhất mô tả diễn biến khối lượng các nhóm thỏ theo thời gian thí nghiệm.
4. Xác định mô hình phân tích phương sai (ANOVA), chạy phần mềm máy tính (SAS hoặc MINITAB) và nộp kết quả phân tích (ghi rõ các câu lệnh và bản copy kết quả phân tích của phần mềm máy tính) cho các chỉ tiêu:
 - Khối lượng đầu kỳ
 - Khối lượng cuối kỳ
 - Tăng khối lượng cả kỳ
 - Tăng khối lượng bình quân
5. Viết và nộp bản thảo bài báo để công bố kết quả trên ở *Tạp chí Khoa học và Phát triển* của Học viện Nông nghiệp Việt Nam.
6. Chuẩn bị báo cáo kết quả thí nghiệm trên bằng PowerPoint và trình bày seminar (nếu có yêu cầu của giảng viên).

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Nguyễn Đình Hiền và Đỗ Đức Lực (2007). Giáo trình thiết kế thí nghiệm. Nhà xuất bản Nông nghiệp, Hà Nội.
2. Nguyễn Văn Tuấn (2012). Từ nghiên cứu đến công bố - Kỹ năng mềm cho nhà khoa học. Nhà xuất bản Tổng hợp thành phố Hồ Chí Minh
3. Vũ Cao Đàm (2005). Phương pháp luận nghiên cứu khoa học. Nhà xuất bản Khoa học và Kỹ thuật. Hà Nội
4. Vũ Đình Hoà và Vũ Thanh Hương (2014). Phương pháp nghiên cứu khoa học. Nhà xuất bản Đại học Nông nghiệp, Hà Nội

Tiếng nước ngoài

1. Anthony C. W. and J. R. McCuenm (1989). *Writing the Research Paper*, Harcourt Brace Jovanovich Publishers.
2. Day R. A. (1998). How to write and publish a scientific paper, 5th edition. Phoenix Oryx Press.
3. Kaps M. and W. R. Lamberson (2004). *Biostatistics for Animal Science*. CABI Publishing.
4. Kothari C. R. (2004). Research methodology – Methods and Techniques, 2nd edition. New Age International Publishers.
5. Marasinghe M. G. and W. J. Kennedy (2008). SAS for Data Analysis: Intermediate Statistical Methods.
6. Moula Nassim, Do Duc Luc, Pham Kim Dang, Frédéric Farnir, Vu Dinh Ton, Dang Vu Binh, Pascal Leroy and Nicolas Antoine-Moussiaux (2011). The ri chicken breed and livelihoods in north vietnam: Characterization and prospects. Journal of Agriculture and Rural Development in the Tropics and Subtropics, 112(1): 57-69.
7. Овсянников А. И. (1976). Основы опытного дела в животноводстве. Колос.
8. Preston T. R. (1995). Tropical animal feeding. A manual for research workers. FAO Animal Production and Health Paper 126.
9. Ramon C. L., J. F. Rudolf, C. S. Philip (1991). SAS System for Linear Models (3rd ed.) SAS Institute Inc.
10. Sing Y. K. (2006). Fundamentals of research methodology and statistics. New Age International Publishers.
11. Tuckman B. W. (1999). Conducting Educational Research. Harcourt Brace College Publishers.
12. Tufte E. R. (2001). The visual display of quantitative information (2nd ed). Cheshire, CT: Graphics Press.

NHÀ XUẤT BẢN ĐẠI HỌC NÔNG NGHIỆP

Trâu Quỳ - Gia Lâm - Hà Nội

Điện thoại: 043. 876. 0325 – 04. 6261. 7649

Email: nxbdhnn@vnua.edu.vn

www.vnua.edu.vn/nxb

Chịu trách nhiệm xuất bản

NGUYỄN QUỐC OÁNH

Biên tập

ĐỖ LÊ ANH

PHẠM DIỆU LINH

Thiết kế bìa

ĐỖ LÊ ANH

Chế bản vi tính

TRẦN THỊ KIM ANH

ISBN: 978-604-924-213-7

NXBDHNN - 2016

In 300 cuốn, khổ 19x27 cm, tại Công ty TNHH MTV NXB Nông nghiệp

Địa chỉ: Số 6 ngõ 167 Phương Mai, Đống Đa, Hà Nội

Số đăng ký kế hoạch xuất bản: 678-2016/CXBIPH/07-01/ĐHNN

Số quyết định xuất bản: 01/QĐ-NXB-HVN ngày 31/3/2016

In xong và nộp lưu chiểu quý II năm 2016