

분산분석

박선일¹ · 오태호*

강원대학교 수의과대학 및 동물의학종합연구소, *경북대학교 수의과대학

(게재승인: 2010년 2월 11일)

The Application of Analysis of Variance (ANOVA)

Son-Il Pak¹ and Tae-Ho Oh*

School of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chuncheon 200-701, Korea

*College of Veterinary Medicine, Kyungpook National University, Daegu, 702-701, Korea

Abstract : Analysis of variance (ANOVA) is a method to analyze the data from the experimental designs comparing two or more groups or treatments at the same time, and is the most effective tool of analyzing more complex data sets with different source of variations. This article describes the logic of ANOVA, the application of the method to the analysis of a simple data set, and the methods available for performing planned or *post hoc* multiple comparisons between the treatments means. In addition, the common misuse of the techniques is also discussed to emphasize that an inappropriate statistical analysis is potentially far more harmful than poorly conducted research. Lastly, an example is given for illustration purposes.

Key words : analysis of variance (ANOVA), multiple comparison, experimental design.

서 론

분산분석(analysis of variance, ANOVA)은 두 개 이상의 처리군 간 평균을 비교하기 위하여 처리군 간 분산과 처리 내 관찰치 간의 분산을 이용하여 분석하는 기법이다. 처리군이 2개 일 때 사용하는 t 검정은 분산분석의 특별한 경우로 두 방법에 의한 계산결과는 동일하다. 분산분석에서 사용하는 용어로 처리(요인) (treatment, factor)와 수준(level)이 있는데 전자는 실험에서 연구자의 관심사항으로 선택된 독립변수이고, 후자는 처리가 취할 수 있는 값이나 조건을 의미한다. 예를 들어 새로 개발된 마취제 'A'의 마취지속 시간(효과)을 3개의 투여농도(조건)에 대하여 평가하는 실험에서 처리는 마취제 'A'가 되며, 수준은 3이다. 마취지속 시간을 반응(response)이라 하며 이는 실험에서의 측정결과를 의미한다. 정의상으로 볼 때 요인이 두개 이상인 실험계획에서 처리는 실제로 실험이 수행되는 요인수준의 조합을 의미하는데 예를 들어 요인이 2개이고 각 요인별로 3개의 수준으로 구성된 실험계획에서는 총 9개의 처리수준 조합이 가능하다.

분산분석은 실험계획에 따라 다양한 형태로 구분되며 가장 간단한 완전확률화계획법(completely randomized design,

CRD)은 실험단위를 각 처리군에 무작위로 할당하여 처리군과 관련된 반응에서 차이가 있는지를 평가하는 것으로 처리 이외의 다른 요인을 고려하지 않는 형태의 실험이다. 실험결과로 얻는 반응을 분류하는 기준(처리군)이 1개이므로 일원 배치계획(one-way ANOVA)이라고 한다. 처리요인의 개수가 2개일 때 이원분산분석법(two-way ANOVA), 3개 일 때 삼원분산분석법(three-way ANOVA)이라고 하며, 이외에도 난괴법(randomized complete block design), 요인계획법(factorial design), 분할구계획법(split-plot design), 교차계획법(cross-over design) 등 다양한 실험계획이 있다(2,3,6,9). 처리효과에 따라 고정효과(fixed effect)와 랜덤효과(random effect) 모형으로 분석할 수도 있다(4). 본 연구에서는 분산분석의 기본적인 원리와 귀무가설이 기각될 때 처리군 간 차이를 비교하는 다중검정(multiple comparison)에 대하여 설명한다.

결 론

분산분석의 원리

예를 들어 두 집단의 평균을 비교하는 실험에서 연구자는 두 군의 개체(실험단위)로부터 어떤 속성(관찰치, x)을 측정하여 두 군간 평균이 다른지에 관심을 갖는다. 이러한 자료에 대해서는 t 검정을 사용하며, 두 실험군의 평균 차이($\bar{x}_1 - \bar{x}_2$)를 두 군의 실험개체를 통합할 때 개체 간 변동 즉 차이의

¹Corresponding author.
E-mail : paksi@kangwon.ac.kr

표준오차(SE_{diff})로 나누어 계산한다. 계산결과 검정통계량이 유의하다면 귀무가설을 기각하고 두 실험군의 평균이 유의하게 다르다는 결론을 얻는다.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE_{diff}}$$

여기에서 분자는 두 처리군의 차이(변동성)이고, 분모는 차이의 표준오차로 각 처리군 내 변동성(variability)에 대한 추정치를 의미한다. 이러한 원리는 처리군이 2개 이상인 분산분석에서도 동일하게 적용되며, 변동성을 표준편차가 아니라 분산을 사용한다.

두개 이상의 처리군을 갖는 실험계획에 t 검정의 반복

예를 들어 대조군(t_0)과 2개의 실험군(t_1, t_2)에 대한 실험에서 연구자는 두 처리군 t_1 과 t_2 가 대조군의 평균과 차이가 없다는 가설과 두 처리군 간 평균이 서로 차이가 없다는 귀무가설에 관심을 갖는다. 이러한 가설을 두 독립표본에 대한 t 검정으로 수행하기 위해서는 t_0 대 t_1 , t_0 대 t_2 , t_1 대 t_2 등 t 검정을 3회 반복해야 한다. 만일 3회의 비교가 모두 독립적으로 수행된 것이 아니라면 문제가 발생한다. 이를테면 $t_0 > t_1$ 이고 $t_0 = t_2$ 이면 $t_2 > t_1$ 이 성립되며 이 경우 t_2 와 t_1 의 비교는 앞의 두 검정 결과로부터 얻은 것이지 독립적으로 검정한 것이 아니다. 이러한 문제를 해결하기 위하여 세 집단의 평균이 동일한지에 대한 동시분석 방법으로 개발된 기법이 분산분석이다. 이는 실제 가설검정에서의 제1종 오류 확률(type I

error)이 지정한 유의수준(참인 귀무가설을 잘못하여 기각할 확률로 일반적으로 5%를 사용) 보다 크게 계산되는 결과를 초래하는 문제점으로도 설명할 수도 있다. 전술한 예에서 3개의 귀무가설을 t 검정으로 반복하여 비교할 때 3개의 비교 중 적어도 1개에서 제1종 오류를 범할 확률은 15%(0.05×3)로 증가하며 이를 제1종 오류가 팽창(inflated) 되었다고 한다. 따라서 유의성을 올바르게 판정하기 위해서는 연구자가 유의수준을 5%로 지정하였다면 다중비교를 위한 제1종 오류 확률이 모든 검정에서 5%가 유지되도록 조절할 필요가 있다. 보다 자세한 내용은 비교오류율에서 설명한다.

변동의 분할

실험군이 k개인 CRD에 대한 분산분석에서 j번째 처리군에서 j번째 관찰치의 총 변동(total variation, total sum of square, SST)은 처리군 간(between treatment)의 변동(between sum of square, SSB)과 처리군 내 개체 간(within treatment)의 변동(오차제곱합) [within sum of square (SSW), error sum of square (SSE)]으로 분할한다(Table 1) (6).

$$SST = SSB + SSE$$

이 자료에 대한 분산분석 표는 Table 2와 같다. 처리군 간 평균 제곱합을 처리군 내 평균 제곱합으로 나누면 분산비(variance ratio, F ratio)를 얻고 계산된 검정통계량을 분자의 자유도($k - 1$)와 분모의 자유도($N - k$)에 대한 임계값과 비교하여 유의성을 판단하게 된다. 만일 k개 처리군의 평균

Table 1. Partition of sum of squares in a one-way ANOVA in a completely randomized design with k treatment groups

Subject	Control(t_0)	Group 1(t_1)	Group 2(t_2)	...	Group k(t_k)	Total
1	x_{10}	x_{11}	x_{12}	...	x_{1k}	
2	x_{20}	x_{21}	x_{22}	...	x_{2k}	
3	x_{30}	x_{31}	x_{32}	...	x_{3k}	
Totals	$\sum x_{.0} = T_{.0}$	$\sum x_{.1} = T_{.1}$	$\sum x_{.2} = T_{.2}$		$\sum x_{.k} = T_{.k}$	$T_{..}$
Means	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$		$\bar{x}_{.k}$	$\bar{x}_{..}$
$T_{..}$	$\sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}$					
SST	$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}^2 - T_{..}^2 / N$					
SSB	$\sum_{j=1}^k n_j (\bar{x}_{.j} - \bar{x}_{..})^2 = \sum_{j=1}^k T_{.j}^2 / n_j - T_{..}^2 / N$					
SSE	$SST - SSB = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij})^2 - \sum_{j=1}^k (T_{.j})^2 / n_j$					

* SST, total sum of square; SSB, between sum of square; SSE, error sum of square; k, number of treatment groups; $T_{..}$, grand total; $\bar{x}_{..}$, grand mean; n, number of observation in each treatment group; N, total number of observation.

Table 2. ANOVA table for a one-way completely randomized design

Source of variation	SS	df	MS	F
Treatments	SSB	$k - 1$	$MSB = SSB / (k - 1)$	MSB / MSE
Error	SSE	$N - k$	$MSE = SSE / (N - k)$	
Total	SST	$N - 1$		

* SS, sum of square; df, degree of freedom; MS, mean square, F, test statistic.

간 유의한 차이가 없다면 N 개의 총 관찰치는 공통 모집단 평균(μ)을 갖고, 처리군 간 분산이나 처리군 내 분산이 동일하다는 것을 의미한다. 반대로 처리군 간 평균제곱이 처리군 내 평균제곱 보다 크면 분산비가 증가할 것이고 이는 처리군 간 평균이 다르기 때문에 귀무가설을 기각한다는 것을 의미한다.

$$F = \frac{SSB/(k-1)}{SSE/(N-k)} = \frac{MSB}{MSE}$$

처리 평균간 비교

고정효과모형(type I model, fixed effect model)에 대한 분산분석에서 처리군 간 차이가 있을 경우 구체적으로 어느 처리군 간 평균이 유의한 차이가 있는지를 분석하게 되는데 이를 다중비교라 한다. 예를 들어 처리군이 3개인 실험에서 귀무가설을 기각하는 경우 이것이 반드시 세 처리군 간 평균이 모두 다르다는 것을 의미하는 것이 아니고 특정 처리군 간 이질이면 t_0 와 t_2 군만 차이가 있고 나머지 군 간에는 차이가 없을 수 있다. 따라서 다중비교에서는 가능한 모든 처리군 간 비교를 수행하게 된다. 이러한 비교는 실험을 수행하기 이전에 설정하여 이미 계획된 비교(planned or priori comparison)를 수행하거나 실험을 수행하여 자료를 얻은 후 모든 처리군 간 상호비교(post hoc comparison)의 목적으로 수행될 수 있다.

사전계획 비교

처리군이 많을 때 연구자는 다양한 비교에 관심을 둘 수 있다. 사전에 계획되지 않은 비교를 수행하는 경우(실제로 검정하지 않지만) 모든 가능한 비교를 고려하기 위해서는 계산된 유의확률(p-value)을 더 작은 값으로 보정해주어야 한다. 이를테면 처리 간 평균이 동일하다는 귀무가설을 유의수준 5%에서 검정할 때 100개의 비교 중 5개는 유의할 것으로 기대할 수 있다. 즉 귀무가설이 참일지라도 총 비교 횟수 중 5%는 유의한 결과를 얻는다는 것을 의미한다. 이러한 5개의 비교가 유의하게 나오는 것은 분명히 잘못된 것이므로 유의수준을 0.05 이하로 설정함으로써 모든 비교에 대한 총 확률이 0.05 이하가 되도록 조절하는 것이다. 따라서 실험을 수행하기 이전에 관심을 두고 있는 비교조건에 대하여 사전계획을

수립하는 것이 바람직하며(흔히 대비를 이용하여 검정함) 이 경우 실제로 검정하지 않은 모든 비교를 고려하기 위하여 계산된 유의확률을 보정할 필요가 없다.

직교 비교

직교대비(orthogonal contrast)는 한 쌍의 선형대비(linear contrast: Q_1, Q_2)에 대하여 처리수준의 효과(c_i : Q_1 에서 i 번째 처리군의 계수; d_i : Q_2 에서 i 번째 처리군의 계수)의 곱($c_i d_i$)의 합($\sum c_i d_i$)이 0일 때 직교한다(독립적이다)는 용어를 사용한다(9). 직교대비 제곱합을 모두 더하면 처리제곱합이 된다는 점에서 중요한 의미를 갖는데 이는 사전계획된 직교비교에서 k 개의 처리군이 있을 때 처리제곱합(분자)의 자유도는 $k-1$ 이 되며 처리 제곱합은 $k-1$ 개의 독립적인 직교대비로 분할할 수 있음을 의미한다(예시 참고). Table 3의(A)에서 3개의 비교(Q_1, Q_2, Q_3)는 어느 쌍에 대하여도 대응되는 계수들을 곱한 후 더하면 항상 0이 되므로 직교대비에 해당하며, 이 때 총 제곱합은 다음과 같이 분할된다.

$$SST = SSB + SSE = SS1 + SS2 + SS3 + SSE$$

마찬가지로(B)에서 제시된 2개의 비교에서 계수 간 곱의 합은 0이므로 직교비교에 해당하지만(C)의 경우 대비 1과 대비 2의 계수 간 곱의 합이 1이 되므로 직교비교가 아니다. 전술하였듯이 사전계획된 다중비교가 직교비교인 경우 유의확률을 보정할 필요가 없는데 직교비교는 모든 비교가 독립적인 경우로 동일한 평균을 2회 비교하지 않기 때문이다.

직교비교에 대한 이해를 돕기 위하여 예를 들어 9개 운동종목 [축구, 농구, 럭비, 수영, 소프트볼, 배구, 라크로스(lacrosse), 조정, 크로스컨트리(cross-country)]에서 활동하고 있는 선수들의 장심(height of the arch of the foot)을 측정하는 실험을 가정하자 (Table 4). 상황 1은 2개의 직교비교이고, 상황 2는 3개의 직교비교에 해당한다. 반면에 상황 3과 4의 경우 배구와 수영은 점핑(jumping) 종목(농구와 배구)과 non-jumping 종목 간 비교에 포함되고(상황 3), 수영과 배구는 대비 Q_1 의 비교에 이미 포함되어 있기 때문에(상황 4) 직교비교가 아니다. 사전계획된 비교를 수행할 때 통합 자료(pooled data)에 대하여 분산분석을 수행할 수 있다. 이를테면 사전계획된 비교로 jumping과 non-jumping 종목 간 비

Table 3. Examples of orthogonal or non-orthogonal contrast of planned comparison with different treatment groups

Comparison	Experiment	Contrast	Sum of square (SS)	Contrast coefficient			
				\bar{x}_1	\bar{x}_2	\bar{x}_3	\bar{x}_4
Orthogonal	A	Q1	SS1	1	-1	0	0
		Q2	SS2	0	0	1	-1
		Q3	SS3	1	1	-1	-1
	B	Q1	SS1	-2	1	1	
		Q2	SS2	0	-1	1	
	Non-orthogonal	C	Q1	SS1	1	0	-1
Q2			SS2	1	-1	0	

Table 4. Comparison of orthogonal and non-orthogonal contrast

Situation	Contrast	FB	BB	RB	SB	VB	LC	SW	CR	CC
1	Q1	■	□							
	Q2							■		□
2	Q1	□	■			■		□		□
	Q2			▲			■			
	Q3				○				▲	
3	Q1	□	■			■		□		□
	Q2					▲		○		
4	Q1	□	□	□	□	□	□	■	■	■
	Q2	○	▲			▲		○		○

*FB = football, BB = basket ball, RB = Rugby, SB = Softball, VB = Volleyball, LC = Lacrosse, SW = Swimming, CR = crew, CC = cross-country. Symbols represent comparisons between different symbols in each contrast.

Table 5. Statistical methods of making post-hoc multiple comparisons

Method	Comparison	Error control	Equal size*	ANOVA assumption	
				Equal variance	Normality
Fisher's LSD	All pairwise	CER	Yes	Yes	Yes
Duncan's MRT	All pairwise	CER	No	Yes	Yes
SNK MRT	All pairwise	EER	Yes	Yes	Yes
Tukey HSD	All pairwise	EER	No	Yes	Yes
Scheffe	All pairwise	EER	No	No	No
Games-Howell	All pairwise	EER	No	No	No
Dunnett	Compare treatments with control	EER	No	No	Yes
Bonferroni (Dunn)	All pairwise and treatments with control	EER	Yes	Yes	Yes

*Equal size=equal number of observations in each treatment group. Abbreviations: LSD=protected least square difference, MRT=multiple range test, SNK=student-Newman-Keuls, HSD=honestly significant difference, CER=comparison-wise error rate control, EER=experiment-wise error rate control

교를 수행하는 경우 농구와 배구종목의 자료를 하나의 처리군으로 통합하고 기타 종목의 자료를 하나의 처리군으로 통합하여 두 처리군에 대하여 분산분석을 수행하고 그 결과로 얻는 유의확률은 보정하지 않고 그대로 사용하면 된다.

사전계획된 비교의 중요성

전술하였듯이 사전계획된 비교는 실험을 수행하기 이전에 수립되어 있어야 한다. 만일 실험을 수행하여 얻은 자료를 검토한 후 특정한 군 간의 비교를 사전계획된 비교인 것처럼 수행하면 이는 비과학적인 접근이다. 예를 들어 Table 4에서 크로스컨트리 종목 선수들의 장심이 가장 낮고 수영 선수들이 가장 높은 것으로 나타난 경우 이들 두 종목 간 장심을 비교하면 유의확률은 매우 낮게 계산된다. 이러한 결과는 9개의 처리군에 대한 총 36개의 가능한 비교 쌍 중 단 하나의 비교만을 수행하였기 때문이다. 모든 처리군의 평균이 실제로 동일하다고 하더라도 유의수준 5%에서 20개 비교 중 1개는 유의할 것으로 기대할 수 있으므로 36개 비교 중 매우 극단적인 두 군을 비교할 때 유의한 결과를 얻는 것은

당연히 예상할 수 있다.

직교다항분석

처리제곱합을 직교대비로 분할하는 다른 방법으로 직교다항분석(orthogonal polynomial analysis)을 수행할 수 있다. 이 분석은 처리제곱합을 처리수준들의 1차, 2차 혹은 3차함수로 설명할 수 있는 변동으로 분해하여 요인수준에 따라 반응값이 보이는 양상을 분석할 수 있다. 이를테면 처리군을 순서로 정렬이 가능할 때 약물의 농도에 따라 각 처리군을 배정한 실험에서 처리군 간 평균에 차이가 있다는 것은 세 군의 평균이 선형효과(일차효과, linear effect), 2차효과(quadratic effect) 혹은 3차효과(cubic effect)로 설명할 수 있음을 의미하며 이를 직교다항분석으로 검정할 수 있다. 직교대비의 개수는 요인의 자유도 즉 처리 수-1이므로 처리군이 5개라면 직교다항식은 1차, 2차, 3차, 4차 등 총 4개의 직교대비로 요인제곱합을 분해할 수 있다. 반응값을 고차원의 효과로 평가하게 되면 결과를 해석하는 것이 쉽지 않기 때문에 1차나 2차효과로 변동을 설명하는 경우가 많다. 이 분석을 위하여

사용되는 계수는 관련서적을 참고하기 바란다(9).

사후비교

실험을 통하여 자료를 얻은 후 처리 간의 평균을 비교하는 사후비교(*post hoc*)용으로 사용되는 기법은 매우 많다 (Table 5). 이들 검정법은 처리 평균 간 유의성을 달성하는데 필요한 차이를 설정하는 기준, 분산분석의 가정을 위반할 때 검정법의 민감도 정도, 가설검정의 오류를 보호하는 수준 등에 차이가 있다. 예를 들어 Table 5에 제시된 기법 중에서 제1종 오류를 범할 가능성은 Fisher's LSD, Duncan, SNK, Tukey-Kramer's HSD, Scheffe 순으로 높다. HSD는 각 처리군 당 표본크기가 동일하거나 매우 다를 때 적용할 수 있으며 SNK는 제1종 오류 보다 제2종 오류의 가능성이 더 높다. Scheffe는 분산분석의 가정을 위반하는 경우에도 사용할 수 있으며(*robust*) 제1종 오류의 가능성을 최소화하는 방법으로 알려져 있다. Games-Howell법은 처리군 간 분산이 이질성을 보이고 정규성을 만족하지 못하면서 처리군 당 표본크기가 다를 때 *robust*하게 적용할 수 있는 방법이다. 한편 Dunnett 방법은 대조군과 나머지 처리군 간 비교할 때 사용하며 처리군 당 표본크기가 동일하거나 혹은 다른 경우에도 적용이 가능하다. Bonferroni (Dunn)법은 처리군이 매우 많은 실험에서 군간 비교하거나 *false positive*인 결과를 줄이고자 할 때 유용하게 사용할 수 있다.

Bonferroni 보정법

유의수준을 보정하는 가장 일반적인 방법으로 유의수준을 비교횟수(*k*)로 나누어 보정한다. 예를 들어 4개의 비교를 수행하는 경우 개별검정에 대한 유의성을 판단하는 기준으로 $0.05/4 = 0.0125$ 를 사용한다. 이 방법은 실험오류율을 보정하는 방법으로 매우 보수적인 결과(즉 실제로 차이가 있지만 차이가 없는 것으로 판정할 가음성 확률)를 초래하므로 가양성 결과가 매우 중요한 문제가 되는 상황에서 적용하는 것이 바람직하다. 이를테면 특정 질병을 치료하고자 새로 개발된 신약의 효과를 기존의 약제와 비교하는 연구에서 부작용 발생율, 치료율, 체내 대사율 등 20가지 항목에 대한 평가결과 신약의 효과가 있다고 판단하여 시판한다고 하자. 기존 약제는 가격은 비싸지만 효과가 충분히 인정되고 부작용도 없는 반면 신약은 가격이 저렴한 장점이 있다고 할 때 경제적인 측면에서 신약을 사용할 수 있을 것이다. 만일 신약의 효과를 판정하는 20개의 평가항목 중 일부 항목에서 가양성 결과(효과가 없음에도 있다고 판정하는 오류)가 있었지만 연구자가 이를 인지하지 못하였다면 소비자로부터 소송을 당하는 극단적인 상황이 초래될 수 있다. 따라서 유의수준 5%에서 평가항목 중 어느 한 항목에서도 가양성 결과가 발생하지 않도록 검정의 전체적인 유의수준을 하향 조정할 필요가 있는 것이다.

비교오류율과 실험오류율

가설검정에서 제1형 오류는 흔히 5%로 설정하고 계산된

유의확률이 5% 보다 작을 때 유의하다는 결론을 내린다. 오류율을 5%로 낮게 설정하는 이유는 처리효과가 있다는 잘못된 결론을 얻은 확률을 줄이고자 하는 연구자의 의지가 반영된 것이라 할 수 있다. 제1형 오류율은 비교오류율(*comparison-wise error rate, α*)과 실험오류율(*experiment-wise error rate, E*)로 구분할 수 있다. 하나의 실험에서 단 1개의 가설을 검정하는 경우 제1형 오류율은 그대로 유지되지만 1개 이상의 가설을 동시에 검정하는 경우 특별한 보정을 사용하지 않으면 제1형 오류를 범할 확률이 매우 증가한다. 예를 들어 처리가 3개인 경우 연구자는 3개의 귀무가설 $t_0 = t_1, t_0 = t_2, t_1 = t_2$ 에 대하여 검정하는데 분산분석에서는 처리군 1쌍 간의 개별비교가 아니라 3개의 가설을 동시에 비교하므로 다중비교라고 한다. 위의 가설에 대하여 *t* 검정을 반복하면 전술하였듯이 제1종 오류가 증가하는 문제가 발생한다. 이를테면 각각의 *t* 검정에서 유의수준을 5%로 설정하는 경우 3개의 귀무가설(처리개수, *t*) 중 어느 하나를 잘못 기각할 확률은 14.3%로 증가한다.

$$E = 1 - (1 - \alpha)^3 = 1 - (1 - 0.05)^3 = 0.143$$

이러한 결과는 *t* 검정을 반복하면 처리군간 평균이 차이가 없음에도 불구하고 차이가 있다는 잘못된 결론을 얻을 확률이 지정한 유의수준 5%가 아니라 14.3%로 매우 증가한다는 것을 의미하며, 이러한 오류율은 처리군의 개수 즉 비교횟수가 증가하면 현저히 증가한다. 따라서 다수의 쌍별 비교를 동시에 검증할 때 제1종 오류율을 조절하고자 다중비교를 사용하는 것이며, 처리군이 2개일 때 비교오류율과 실험오류율은 동일하다.

실험오류율과 비교오류율 중 어느 것이 더 중요할지를 판단하는 기준은 없고 연구상황에 따라 연구자가 주관적으로 판단한다. 판단기준에서 첫째, 실험오류율을 조절한다는 것은 제1종 오류(참인 귀무가설을 기각하여 차이가 있다고 판단하는 오류)는 줄이는 반면 제2종 오류(거짓 귀무가설을 수용하는 오류)는 증가할 수 있으며 둘째, 동일한 실험에서 한 가지 추론을 잘못하는 것이 나머지 추론에 영향을 미치지 못한다면 비교오류율을 사용할 수 있다는 점을 고려할 필요가 있다. 일반적으로 실험을 수행하기 이전에 가설을 설정한 경우에는 비교오류율을 조절하고 실험 종료 후 사후분석에서는 실험오류율을 조절한다.

분산분석의 가정 검토

적절하지 못한 통계분석을 사용하면 연구결과를 잘못 해석하고 결국 왜곡된 결론에 도달 할 수 있다는 점에서 실험을 수행하는 과정에서 범한 오류 보다 더 심각한 문제를 초래할 수 있다. 2006-2007년 기간 동안 한국임상수의학회지에 발표된 129 편중 1개 이상의 통계분석을 수행한 94편의 논문 중 62편(66%)의 논문에서 적어도 하나의 통계적 오류가 발견되었다. 분산분석과 관련하여 자료에 대한 가정 검토, 다중비교 방법 선택, 반복측정 자료에 대한 의존성 가정, 통계 분석 방법에 대한 설명, 분석기법에 대한 구체적인 설명 등

Table 6. Data summary (weight in kg)

Subjects	Group 1	Group 2	Group 3	Group 4
1	62.3	78.2	79.3	78.8
2	54.2	69.1	67.5	68.5
3	48.7	77.0	55.8	86.3
4	69.4	68.1	69.0	74.9
5	77.8	69.4	73.6	86.7
6	69.3	88.3	55.8	80.3
Sample mean	63.6	75.0	66.8	79.3
Sample SD	10.8	7.8	9.5	6.9

*SD, standard deviation.

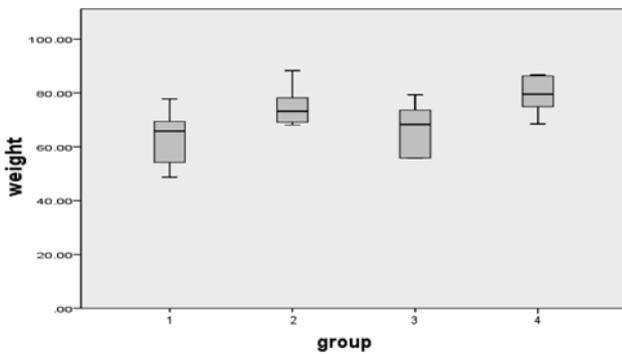


Fig 1. Box plots of the weight (kg) in pigs for each type of diet.

이 문제점으로 나타났다. 분산분석에서 관찰치는 관찰치의 총 평균, 처리 간 변동 및 정규분포 모집단으로부터 추출된 무작위 변동 등 세가지 성분의 합으로 이루어진다. 분산분석에서는 첫째, 오차(무작위 변동)는 평균이 0이고 표준편차가 1인 정규분포를 만족할 것(잔차의 정규성) 둘째, 각 처리군의 평균은 다를지라도 분산은 모든 처리군에서 동일할 것(등분산성) 셋째, 각 실험단위가 상호 독립일 것 넷째, 모형의 각 항(term)은 가법(additive)으로 구성된다고 가정한다. 이러한 가정을 위반하면 유의수준과 가설검정의 민감도에 영향을 미치므로 가정을 충족하지 못할 때 자료변환이나 비모수분석 혹은 분산의 이질성을 가정한 Welch 검정 등을 고려해야 한다(1). 잔차(residual)는 관찰치와 반응모형에서 추정된 예측치의 차이로 잔차에 대한 정규확률도표(normal probability plot of residual)를 작성하여 검토할 수 있다. 처리군 간 분산이 동일하면 잔차는 고르게 분포하며, 잔차의 정규성이 충족되면 정규확률도표에서 직선의 형태를 취한다. 이러한 등분산성에 대한 통계적 검정으로는 Levene's test, Bartlett's test 등을 사용한다.

예 시

자돈(piglet)의 사료로 시판 중인 4종류 사료가 돼지 증체에 차이를 보이는지에 대한 실험을 수행하였다. 총 24두의

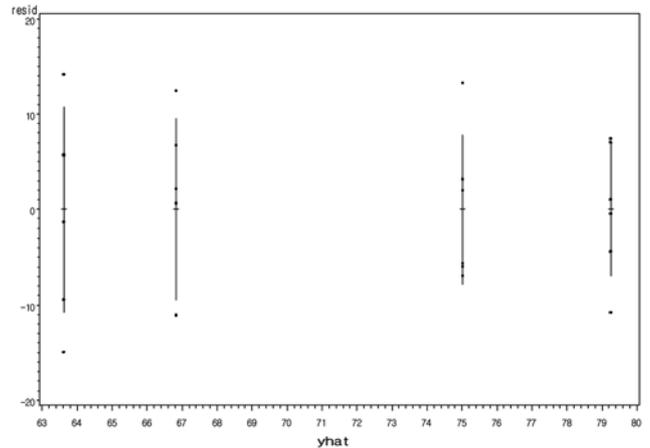


Fig 2. Plot of residuals versus fits for the data in Table 6. Fitted values (yhat) represent treatment means by diet and response is weight in kg.

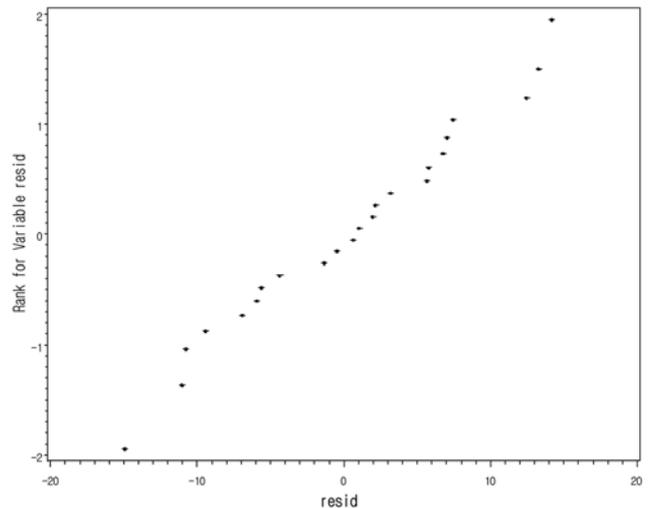


Fig 3. Normal probability plot for the data in Table 6. Resid represent residual and response is weight in kg.

돼지를 각 사료에 6마리씩 무작위로 할당하고 모든 조건을 동일한 상태로 유지하여 실험 종료 후 체중(kg)을 측정한 결과가 Table 6과 같다고 할 때 사료 종류에 따라 증체효과에 차이가 있는지 검정하여 보자.

본 자료에 대하여 사료 종류별 기술통계량을 요약하면 Fig 1과 같다. 먼저 분산분석에 대한 가정을 만족하는지에 대하여 Table 6의 자료에 대하여 잔차 도표를 작성한 결과 잔차가 적합치에 대하여 고르게 분포하고 있어(Fig 2) 등분산성을 만족하는 것으로 파악되며, 통계검정에서도(Levene's test) $p = 0.5092$ 로 계산되어 모든 처리군에서 분산이 동일하다는 귀무가설을 기각하지 못한다는 것을 알 수 있다(Table 7). 또한 잔차에 대한 정규확률도표에서도 직선 형태를 보이고 있어 정규성을 만족하는 것으로 판단된다(Fig 3). 총제곱합을 처리 간 변동과 처리 내 변동으로 분할하고 각각의 자유도로 보정하면 평균제곱을 얻고 두 변동간의 비는 $F = 3.96$

Table 7. ANOVA table for the data in Table 6

Source of variation	SS	df	MS	F	p-value
Treatment (diet)	935.7	3	311.9	3.96	0.023
Error (within diet)	1575.5	20	78.8		
Total	2511.2	23			

* SS, sum of square; df, degree of freedom; MS, mean square, F, test statistic.
Homogeneity of variance test (Levene's test): p=0.5092

Table 8. ANOVA table for the data in Table 6 using the three planned comparisons

Source of variation	SS	df	MS	F	p-value
Treatment (diet)	935.7	3	311.9	3.96	0.023
Q1	457.6	1	457.6	5.81	0.026
Q2	462.5	1	462.5	5.87	0.025
Q3	15.6	1	15.6	0.20	0.661
Error (within diet)	1575.5	20	78.8		
Total	2511.2	23			

* SS, sum of square; df, degree of freedom; MS, mean square, F, test statistic.

(p = 0.023)으로 계산되며 사료의 종류에 따라 체중이 다르다는 결론을 얻는다(Table 7).

실험을 수행하기 이전에 처리군 간 비교에 대한 사전계획이 없었다고 가정하여 보자. 분산분석에서 사료종류에 따라 체중의 효과가 다르다는 결론을 얻었기 때문에 구체적으로 어느 사료 간에 차이가 있는지 알아보고자 Duncan 다중범위 검정을 시행한 결과 유의수준 5%에서 1군과 2군, 1군과 4군, 3군과 4군간 유의한 차이가 있는 것으로 분석되었다.

한편 연구자가 사전계획으로 group 1(대조군 사료)과 나머지 3개 사료(신규 개발 사료)의 평균과 동일한지에 대하여 관심을 둔다고 가정하자. 이는 전술한 대비를 이용하여 분석할 수 있다. 각 처리군의 평균을 $\mu_1, \mu_2, \mu_3, \mu_4$ 라 하면 3, -1, -1, -1의 직교 대비계수(contrast coefficient)를 얻고 이를 귀무가설에 투입한다.

$$\text{직교대비 1: } \mu_1 = \frac{(\mu_2 + \mu_3 + \mu_4)}{3} \Leftrightarrow 3\mu_1 - \mu_2 - \mu_3 - \mu_4 = 0$$

본 예에서 처리제곱합의 자유도가 3이므로 전술하였듯이 처리제곱합은 서로 독립인 3개의 직교대비로 분할할 수 있다. 이를테면 3군과 4군간 비교(직교대비 2), 1군을 제외하고 2군과 나머지 두 군(3군과 4군)의 평균 간 비교(직교대비 3)에 관심을 둘 때 직교대비 계수는 직교대비 2의 경우 0, 0, 1, -1, 직교대비 3은 0, -2, 1, 1이 된다.

$$\text{직교대비 2: } \mu_3 = \mu_4 \Leftrightarrow 0\mu_1 - 0\mu_2 + \mu_3 - \mu_4 = 0$$

$$\text{직교대비 3: } \mu_2 = \frac{(\mu_3 + \mu_4)}{2} \Leftrightarrow 0\mu_1 + 2\mu_2 - \mu_3 - \mu_4 = 0$$

이상의 대비를 사용하여 분석한 결과는 Table 8과 같다.

Table 7에서 처리제곱합은 935.7이었고 이를 3개의 직교대비로 분할한 결과 직교대비 1(457.6)과 2(462.5)는 유의하였으나 직교대비 3(15.6)은 유의하지 않은 것으로 분석된다(Table 8). 한편 앞에서 처리제곱합을 직교대비로 분할하는 다른 방법으로 직교다항분석을 설명하였는데 본 자료의 경우 처리군이 명목형으로 순서정렬이 불가능하기 때문에 이 분석은 수행할 수 없다.

요 약

분산분석은 t 검정을 한 개의 이산형 독립변수가 2개 이상의 수준을 갖는 자료로 확장한 검정이다. 따라서 일원분산분석은 모수 t 검정의 가정이 그대로 적용되며, 비교 집단이 2개인 경우 분포 간의 관계 $F = t^2$ 의 관계에 의해 동일한 결과를 얻는다. 귀무가설은 모집단 평균 간 차이가 없다는 것이고, 대립가설은 적어도 하나의 평균은 다르다고 설정한다. 일원분산분석에 대응하는 비모수적 검정으로 Kruskal-Wallis 검정을 사용하며 자료의 척도가 적어도 순위형이어야 한다 (5). 독립변수가 두개 이상인 이원분산분석의 경우 독립표본, 자료의 정규분포, 처리군 내 분산의 동질성, 연속형 자료에 대한 가정을 요구한다. 이원분산분석에 대응하는 비모수적 검정으로 Friedman 검정이 사용되며 자료의 척도가 적어도 순위형이어야 한다. 한편 짝지은 t 검정을 확장한 연구가 가능하다(7,8). 예를 들어 실험군이 2개 이상이고 동일한 개체에 2개 이상의 처리를 할당하거나 시간을 달리하여 측정된 반복측정 실험에서는 동일한 개체를 여러 가지 처리 혹은 시간에 대하여 반복적으로 측정하였기 때문에 표본 간 독립성이 결여되므로 이원 반복측정 분산분석(2-way repeated NAOVA)으로 분석한다.

감사의 글

본 연구는 강원대학교 동물의학종합연구소의 지원에 의해 이루어졌으며 이에 감사드립니다.

참 고 문 헌

1. 박선일, 이영원. 자료분석의 기초. 한국임상수의학회지 2009; 26: 189-199.
2. Altman DG. Practical statistics for medical research. London, England: Chapman & Hall, CRC, 1997.
3. Armitage P, Berry G. Statistical Methods in Medical Research, 3rd ed. Blackwell Scientific Publications, Oxford, 1994.
4. Armstrong RA, Eperjesi F, Gilmartin B. The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmol Physiol Opt* 2002; 22: 248-256.
5. Daniel WW. Applied nonparametric statistics. 2nd ed. Thompson Information Publishing Group, Boston, 1990.
6. Glantz SA, Slinker BK. Primer of applied regression and analysis of variance. McGraw-Hill, New York, 1990.
7. Kristensen M, Hansen T. Statistical analyses of repeated measures in physiological research: a tutorial. *Adv Physiol Educ* 2004; 28: 2-14.
8. Ludbrook J. Repeated measurements and multiple comparisons in cardiovascular research. *Cardiovasc Res* 1994; 28: 303-311.
9. Montgomery DC. Design and analysis of experiments. 4th ed. John Wiley & Sons, NY. 1997: 681.