

MS 엑셀 프로그램의 통계분석결과 신뢰성 검증 및 매크로 보완 (분산분석 메뉴를 중심으로)

Test for reliability of MS Excel statistical analysis output and modification of macros (Focused on an Analysis of Variance menu)

김 숙 영(Sook Young Kim)¹⁾

요약

초강력 스프레드시트 기능의 엑셀 프로그램에서 자료를 통계적으로 분석하는 데이터분석 메뉴는 엑셀 2000 이후 수정 보완 되고 있지 않으며 그 활용도도 매우 저조하다.

본 연구에서는 엑셀 통계분석 메뉴 활용도 향상을 위하여 사용빈도가 높은 분산분석의 엑셀 메뉴 실행 결과와 계산식에 의한 이론 결과를 비교하고 고차원 분석 매크로를 개발하였다.

실자료를 엑셀 데이터분석 메뉴를 적용하여 얻은 결과와 계산식에 의한 이론 결과는 요인이 한개인 일원배치법이나 두개인 이원배치법에서 모두 정확히 일치하였다. 즉 엑셀 데이터분석 메뉴의 신뢰성이 검증되었다.

고차원적인 분산분석법인 블록화된 분산분석 결과를 제공하는 매크로 및 일원분산분석법에서 평균 차이 있는 그룹들을 찾아내는 다중비교 분석방법 매크로를 엑셀 함수를 이용하여 개발하였다.

Abstract

Statistical analysis menus of MS Excel software, with powerful spreadsheet functions has not been modified since Excel 2000 Edition and its utilization is very low. To improve utilization of Excel menu for statistical analysis, this research compared outputs of Excel statistical menus and computed test statistics, and developed high-level macros.

Outputs of Excel menus, both oneway layout and twoway layout, on real data are exactly same as the computed test statistics, and therefore, Excel menus for statistical analysis are reliable.

Macros to provide results for Analysis of Variance with a block and multiple comparison of means are developed using Excel functions.

Key words: Analysis of Variance, Excel functions, Statistics

논문 접수 : 2008. 11. 15.

심사 완료 : 2008. 12. 11.

1) 정회원 : 안산공과대학 컴퓨터 정보과

1. 서론

현대의 디지털 사회에서는 모든 분야에서 논리적으로 입증될 수 있는 정보를 수집하고 확실적인 결론에 의하여 분석하는 과학적인 연구 방법론을 적용하여 일반화되고 타당성 있는 결론을 도출하여야 한다. 연구 목적에 따라 설정된 가설들에 대하여 분석 결과가 어떤 가설을 뒷받침 하고 있는가를 확률적으로 분석하는 학문이 통계학이다.

정보화 시대에서 점차 연구 대상이 광범위하게 확대되는 현실에서 모집단 전체를 조사할 수 없으므로 표본 조사가 불가피하다. 일반적으로 표본에서 얻어진 정보는 모집단 특성에 대하여 완전 정보를 가지고 있지 않으므로 모집단 특성의 추정 시에는 불가피하게 오차가 발생한다. 이와같은 오차의 허용 범위를 설정하여 모집단 특성에 대한 일반성을 찾아내고 연구 목적에 부합되도록 설정한 가설들을 확률적 기법으로 분석한다. [1,2]

통계 모형을 설정하고 수집된 정보를 통계량에 적용하여 확실값을 계산하는 과정에서 컴퓨터를 사용하게 된다. 1985년 최초의 통합형표 계산용으로 개발된 테이터를 행과 열 개념을 적용하여 입력하고 셀이라는 최소 단위를 번지로 나타내고 함수를 이용하여 자동계산하고 사용자 요구대로 여러 가지 차트를 만들어 주는 스프레드 시트의 가장 막강한 응용프로그램인 EXCEL 이 개발되었다.

단순한 시트 기능 이외에 계속 성능이 향상되어 그래픽 환경, 매크로, DB 기능과 지도차트 작성 등 통합 문서 작성에 필요한 기능도 제공하는 통합형 표 계산 S/W 로 발전되어 기업 업무에서는 필수 S/W 로 발전되었다. 가장 최근 버전인 엑셀 2007은 작업 영역에서 행과 열 증가로 하나의 시트에서 관리할 수 있는 데이터 관리 능력 향상, 데이터 정렬이 64 단계까지 확대, 중복 데이터 제거 기능들이 추가되었다. [3,4]

통계 모형에 따른 통계량을 계산하고 자료 분포도에 따라 확실값을 계산하여야 하는 추측 통계학에서는 통계량을 계산하는 부 프로그램들로 구성된 통계 패키지 프로그램인 Minitab,

BMDP, SPSS, SAS 등이 사용되고 있다.

이러한 통계 패키지 프로그램을 사용하기 위하여는 데이터 파일을 별도로 작성하고 각 프로그램 코드를 작성하기 위한 언어를 터득해야 한다. 또한 전문적인 통계 결과들이 출력되므로 전문가 도움이 필요하다.

현재 엑셀에도 추측 통계학 분석 도구로 사용될 수 있는 데이터 분석 메뉴가 포함되었다. 아직은 기본 자료 분석 방법인 분산 분석과 회귀 분석 결과만을 얻을 수 있다.

가설 검정이 포함된 자료 분석 프로그램을 C 언어 등의 프로그램으로 개발하기 어렵고 전문 통계 패키지를 사용해야하는 중요한 이유는 복잡한 함수에 의하여 계산되는 분포도 값 때문이다.

엑셀에는 통계 분포 함수 및 복잡한 연산 함수 계산이 가능하므로 추측 통계학 가설검정 결과를 제공하는 통계 분석 매크로를 개발할 수 있다. 연구에서 수집된 다량의 정보들은 별도로 자료 파일로 저장되며 통계 패키지 사용 시 계산된 프로그램 자료 파일을 부르는 과정, 또한 자료를 부분만 필요로하는 등의 자료를 관리해야 할 필요가 없다.

본 연구에서는 엑셀에 자료 분석 기능 활성화 향상을 위하여 현재 데이터 분석 메뉴에 있는 분산 분석 매크로 결과의 중요성을 분석하고 완전한 분산 분석 결과를 제공하도록 수정 보완 함이다.

2. 조사방법

2-1. 이론적 배경

2-1-1. 분산분석법 정의 및 가설

Analysis of Variance는 요인 효과를 검정하는 주로 실험 연구에서 적용되는 분석방법이다. 실험 대상은 두 개 이상의 그룹으로 분류하고 각 그룹에 요인들을 다양하게 적용한 후 결과를 측정하고 그룹간의 차이성 유무를 검정한다.

모든 그룹간의 측정 평균이 동일하다면 요인 효과는 없다. 반면에 그룹간의 측정치 평균이 모두 동일하지 않다면 요인 효과가 있는 것이

다.
 요인 효과가 있는 경우 평균이 서로 다른 그룹을 찾아내는 다중 비교 까지가 분산분석에 포함된다. 두 개 이상 그룹의 평균을 비교하는 것 이지만, 그룹과 그룹내의 분산비로 통계량이 추정되므로 분산 분석이라 한다. [5]
 요인이 한 개인 경우는 일원 분산 분석, 두 개인 경우는 이원 분산 분석이라 한다.
 요인이 두 개인 경우는 각 요인간이 동시에 작용하는 교호작용도 생각해야 한다.

2-1-2. 통계모형

추측 통계에서 수집된 표본 정보로 모집단 모수를 추정하기 위하여는 통계 모형을 설정해야 한다.

일원 분산 분석법의 통계 모형은

$$Y_{ijk} = \mu + \alpha_i + \epsilon_{ij}$$

이원 분산 분석법의 통계 모형은

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$$

2-1-3. 가설

분산분석법의 연구목적은 요인이 결과에 미치는 영향을 검정함 이므로 연구 가설의 영가설은 요인의 효과가 없다 와 대립가설은 효과가 있다 이다.

통계 가설은 각 그룹마다의 평균이 동일하다 와 모든 그룹의 평균이 동일하지 않다 이다.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_n$$

$$H_1 : \text{not } H_0$$

2-1-4. 통계량 추정

요인 효과를 통계적으로 검정하기 위하여 그룹 변수 (Treatment) 를 설정하고

그룹간 분산과 그룹내의 분산비로 검정통계량이 산출된다. [6]

2-1-4-1. 일원분산분석법 (Oneway Analysis of Variance)

요인이 한 개인 분산분석법 이므로 k 개 그룹으로 구성된다는 가정하에 자료는 X_{ij} (j: 그룹 변수, $j=1,2,\dots,k$)로 표시된다.

각 그룹내의 측정치 개수 (k)는 반드시 동일할

필요가 없다.

그룹간 분산은 그룹 평균과 총 평균 편차의 제곱합으로 정의되므로 계산식은

$$\sum_{j=1}^k n_j (\bar{X}_{.j} - \bar{X}_{..})^2 \text{ 이고 간편 계산식은}$$

$$\sum_{j=1}^k \frac{T_j^2}{n_j} - \frac{T_{..}^2}{N} \quad (T_{..} : \text{자료 총 합, } N: \text{자료 총 개수}) \text{ 이다.}$$

그룹내 분산은 그룹내에서 측정치와 그룹평균 편차의 제곱합으로 정의되므로 계산식은

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{.j})^2 \text{ 이고 간편 계산식은}$$

$$\sum_{j=1}^k \sum_{i=1}^{n_j} X_{ij}^2 - \sum_{j=1}^k \frac{(T_j)^2}{N} \quad (n_j : \text{각 그룹별 측정치개수, } T_j: \text{각 그룹별 측정치합}) \text{ 이다.}$$

모든 가설검정에서 결론을 위한 최종 추정치는 p 값이며 분산분석법에서의 p 값은 산출된 검정통계량 값이 F 확률 분포도 (자유도가 k-1 과 N-k) 값보다 크거나 같을 확률이다. 즉 $p = \text{Probability}(VR \geq F_{k-1, N-k})$ 와 같이 표현될 수 있다.

연구자가 정해놓은 오차 허용범위인 유의도 (a) 값보다 적을 때 ($p < a$) 일때는 영가설이 기각되어 즉, 그룹변수의 평균값이 모두 동일하지 않으므로 요인 효과가 있다는 결론을 내린다. 반면 $p \geq a$ 일때는 영가설을 채택하여 모든 그룹의 평균값이 동일하므로 요인 효과가 없다는 결론을 내린다 (표 1).

<표 1> 일원분산분석표 구성

요 인	자승합	자유도	평균자승합 (자승합/자유도)	검정통계량
그룹간	SSA	k-1	MSA	MSA/MSW
그룹내	SSW	N-k	MSW	

2-1-4-2. 블록화된 일원 분산 분석법

각 그룹내에서 블록이라는 요소를 적용하여 그룹내 분산인 오차를 감소시키는 분산분석법으로 자료는 X_{ij} (i: 블록, j: 그룹요인) 의 2

차원 배열로 표시된다.

각 그룹내의 측정치 개수 (n) 는 모두 동일하여 I=1,2,...,n, j=1,2,...,k 의 값을 가진다 (표 2).

<표 2> 블록화된 분산분석표 구성

요인	총자승합	자유도	평균자승합 총자승합/자유도	분산비
그룹	SSTr	k-1	MSTr	VR=MSTr/MSE
블럭	SSBI	n-1	MSBI	
오차	SSE	(n-1)(k-1)	MSE	

A요인	B 요인			총합
	1		b	
1	X ₁₁₁ X _{11n}		X _{1b1} X _{1bn}	T _{1..}
a	X _{a11} X _{a1n}		X _{abl} X _{abn}	T _{a..}
총 합	T _{.1.}		T _{.b}	T _{...}

C: 수정계수 $(\sum_{j=1}^k \sum_{i=1}^n X_{ij})^2 / kn$

$$SSTr = \sum_{j=1}^k \frac{T_j^2}{n} - C$$

$$SSBI = \sum_{i=1}^n \frac{T_i^2}{k} - C$$

$$SSE = \sum_{j=1}^k \sum_{i=1}^n X_{ij}^2 - C - SSBI - SSTr$$

최종 검정통계량인 p 값은 $p = \text{Prob}(VR \geq F_{k-1, (n-1)(k-1)})$ 로 계산되며 일원분산 분석법과 동일한 결정원칙을 따른다.

2-1-4-3. 2원 분산분석법

연구 대상인 서로 다른 요인이 두 개일 때 각 요인 효과를 검정하는 분산분석법이다.

각 요인에 대하여 반복적으로 측정하는 경우 세 개의 첨자를 필요로 하므로 (표 3) 자료 배열 형태를 기술하면 자료는 X_{ijk} (i: 요인 A 그룹, j: 요인 B 의 그룹) 의 세 개의 첨자들을 이용하여 표시할 수 있으며, 모든 그룹의 측정 개수는 동일하다 가정한다 (k=1,2,...,n) (표 4).

<표 3> 이원분산분석법 자료 배열

<표 4> 이원 분산분석표

요인	자승합	자유도	평균자승합 (총자승합/자유도)	검정통계량
요인A	SSA	a-1	MSA	MSA/MSE
요인B	SSB	b-1	MSB	MSB/MSE
교호 작용	SSAB	(a-1)(b-1)	MSAB	MSAB/MS E
오차	SSE	ab(n-1)	MSE	

수정계수 $C = (\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk})^2 / abn$

총 자승합: $\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - C$

A 요인자승합: $\frac{\sum_{i=1}^a T_{i..}^2}{bn} - C$

B 요인자승합: $\frac{\sum_{j=1}^b T_{.j}^2}{an} - C$

A, B 동시 효과 요인 자승합:

$$\frac{\sum_{i=1}^a \sum_{j=1}^b T_{ij.}^2}{n} - SSA - SSB - C$$

오차자승합: $\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2$

$$\frac{\sum_{i=1}^a \sum_{j=1}^b T_{ij}^2}{n}$$

2-1-4-4. 그룹간 비교

분산분석 결과 요인 효과가 있을 때 (그룹별 평균이 모두 동일하지 않다) 평균이 다른 그룹들을 찾아야한다. 통계상으로 가장 신뢰성이 높은 결과를 제공하는 방법은 터키 방법 (Tukey's test) 이나, 특별한 참조 표가 필요하므로 t 분포도를 이용한 Bonferoni 방법을 적용한다.

검정통계량 계산식:

$$\frac{|\bar{X}_i - \bar{X}_j|}{\sqrt{\frac{s_1^2}{n_i} + \frac{s_j^2}{n_j}}}$$

p=Prob(검정통계량 > t (자유도 N-k, 2) 분포표

2-2. 분산분석 결과를 얻기 위한 엑셀 메뉴

엑셀에서 분산분석 자료 분석을 수행할 수 있는 메뉴는 일원분산배치법, 반복없는 이원배치법 및 반복있는 이원배치법이다. 일원분산분석 결과를 얻을 수 있는 일원배치법 출력 결과는 그룹에 대한 기술통계량 및 분산분석 (제공합, 자유도, 평균제공합, F비, p 값 및 F 기각치)를 포함한다.

반복없는 이원배치법은 두 요인변수의 각 그룹별 기술통계량 및 분산분석표 (제공합, 자유도, 평균제공합, F비, p값 및 F 기각치) 가 출력된다.

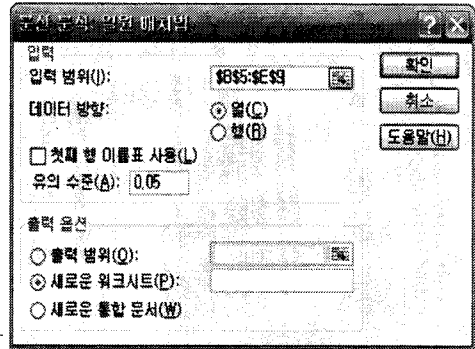
반복있는 이원배치법에서는 두 요인변수들의 그룹들 조합으로 이루어진 모든 쌍 들에 대한 기술통계량들이 출력된다. 분산분석표에는 두 개 각각의 단독 및 두 요인들의 교호작용에 관한 분산분석 결과가 출력된다.

2-2-1. 메뉴 사용법

2-2-1-1. 일원배치법

요인변수 각 그룹의 값들을 엑셀 시트의 행별 또는 열별로 입력한다.

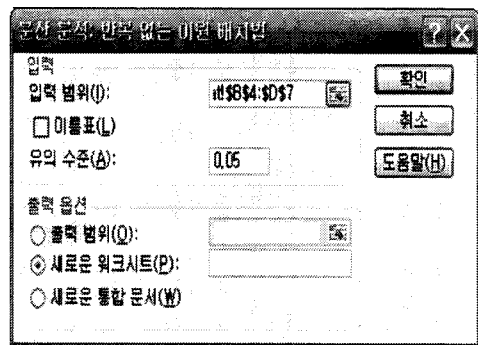
일원배치법 메뉴에서 자료값들이 있는 셀범위, 그룹이 입력된 방향인 데이터방향을 선택하고 오차 허용범위인 유의수준을 입력한다.



<그림 1> 일원배치 메뉴

2-2-1-2. 반복없는 이원배치법

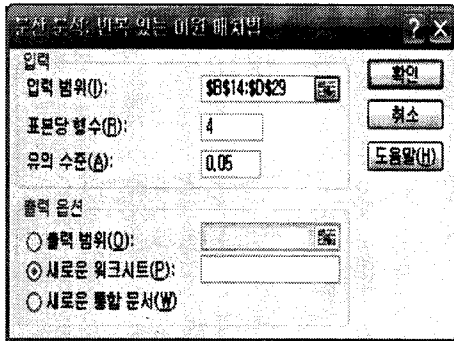
두 개의 요인 변수 중 한 개는 행 방향으로 다른 한 개는 열 방향으로 자료 값들을 입력한다. 메뉴 대화상자에서 입력범위와 유의수준을 입력한다.



<그림 2> 반복없는 이원배치 메뉴

2-1-1-3. 반복있는 이원배치법

두개의 요인변수들 조합으로 생성된 모든 그룹 내에는 값이 두 개 이상이므로 한 그룹에서 데이터 개수인 표본당 행의 수를 입력한다.



<그림 3> 반복있는 이원배치 메뉴

2-3.검정통계량 계산을 위한 코드 작성 엑셀 데이터 분석 메뉴 실행 결과의 정확성을 분석하기 위하여 검정 통계량 계산공식을 수식함수를 적용한 명령코드를 작성하여 직접 계산을 수행하였다.

2-3-1. 일원분산분석 결과 명령 코드

변수내용	코드
그룹간 자승합 (SSW)	=SUMSQ((SUM())/count()) -SUM(SUM())/SUM(COUNT())
그룹내 자승합	=SUMSQ(SUM()) -SUMSQ(SUM())/SUM(COUNT())
p 값	FDIST()
기각역	FINV()

2-3-2. 반복없는 이원분산분석법

변수	코드
A 요인	=SUMSQ((SUM())/count())
자승합	-POWER(SUM(),2)/(count()*COUNT())
B 요인	=SUMSQ((SUM())/count())
자승합	-POWER(SUM(),2)/(count()*COUNT())
총자승합	=SUMSQ()- POWER(SUM(),2) /(count()*COUNT())
p 값	FDIST()
기각역	FINV()

2-3-3. 반복있는 이원분산분석법

변수	코드
A 요인	=SUMSQ((SUM())/count())
자승합	-POWER(SUM(),2)/(count()*COUNT())
B 요인	=SUMSQ((SUM())/count())
자승합	-POWER(SUM(),2)/(count()*COUNT())
요인	=SUMSQ((SUM())/count())
자승합	-POWER(SUM(),2)/(count()*COUNT())
총자승합	=SUMSQ() -POWER(SUM(),2)/(count()*COUNT())
p 값	FDIST()
기각역	FINV()

2-4. 매크로 개발

엑셀 데이터 분석메뉴에는 존재하지 않는 중요 분산분석 계산 매크로를 개발하였다. 블록화된 일원분산분석 및 모든 그룹의 평균이 동일하지 않다는 가설이 기각될 때 차이있는 그룹을 찾는 다중 비교 분석 매크로가 개발되었다.

2-4-1. 블록화된 일원분산분석법 매크로 개발

<표 5> 블록화된 일원분산분석법 매크로 실행 엑셀 시트

	A	B	C	D	E	F	G	H	I
1		1호선	2호선	3호선	4호선		column		
2	1월	2.61	3.15	2.42	2.91			11.07	
3	2월	2.6	3.27	2.49	3.03			11.58	
4	3월	2.46	2.67	2.31	2.71			10.37	
5	4월	2.34	2.62	2.05	2.44			9.45	
6	5월	2.35	2.5	2.04	2.4			9.29	
7									
8	mean	2.516	2.978	2.28	2.699				
9	n	5	5	5	5				
10	rowsum	12.58	14.39	11.3	13.49				
11	공간계산값	31.05128	41.41442	25.536	36.98602				
12	sd	0.03733	0.10967	0.04225	0.07757				
13									
14	n		20 k		4 n				
15									
16	C	133.9549							
17	SST	2.10052							
18	SSti	1.044684							
19	SSB	0.99132							
20	SSE	0.06436							
21									
22	Source	SS	df	MS	F.R.		Pvalue		
23	Tot	1.044684	3	0.348228	64.93723		1.06097E-07		
24	Between	0.99132	4	0.24783					
25	Error	0.06436	12	0.005363					
26									

블록화된 분산분석표 계산 이론에 의하여 주요 셀에 들어갈 함수코드는

계산항목	셀	명령코드
수정계수	B16	=POWER(SUM(B2:E6),2)/B14
총제곱합	B17	=SUMSQ(B2:E6)-B16
그룹요인제곱합	B18	=SUMSQ(B10:E10)/F14-B16
블록제곱합	B19	=SUMSQ(G2:G6)/D14-B16
오차제곱합	B20	=B17-B18-B19
검정통계량	E23	=D23/D25
p 값	E24	=FDIST(E23,C23,C25)

2-4-2. 다중 비교 매크로 개발

분산분석법에서 요인의 효과가 증명될 때 차이가 있는 그룹을 찾아내는 통계 계산 매크로 작성이다.

변수내용	코드
검정통계량	$= (\text{mean}() - \text{mean}()) / (\text{SQRT}(\text{SUMSQ}() - \text{count}() * \text{power}(\text{mean}(), 2))) / ((\text{count}() - 1) / \text{count}())$
p 값	=TDIST()

3. 결과

3-1. 일원분산분석법 실행 결과

4개의 그룹(각 그룹당 자료수는 5개)을 가진 요인 변수에 관한 일원분산분석 결과이다.

<표 6> 일원배치법 메뉴 실행결과

분산 분석: 일원 배치법						
요약표	관측수	합	평균	분산		
인자의 수준						
Column 1	5	12.58	2.516	0.03733		
Column 2	5	14.39	2.878	0.10667		
Column 3	5	11.3	2.26	0.04225		
Column 4	5	13.49	2.698	0.07767		
분산 분석						
변동의 요인	제곱합	자유도	제곱평균	F비	P-값	F기각치
처리	1.04484	3	0.34828	5.278569	0.010098	3.238872
잔차	1.05568	16	0.06598			
계	2.10052	19				

<표 7> 일원분산분석 이론 계산 결과

mean	2.516	2.878	2.26	2.698						
n	5	5	5	5						
rowsum	12.58	14.39	11.3	13.49	sum	51.76	k	4	N	20
중간계산값	31.85128	41.41442	25.538	35.98802						
s2	0.03733	0.10667	0.04225	0.07767						
Source										
	SS	df	MS	Variance Ratio	f	기각치				
Among	1.04484	3	0.34828	5.278569	0.010098	3.238872				
Between	1.05568	16	0.06598							
Total	2.10052									

표 6 과 표 7 에서 보듯이 일원배치법 엑셀메뉴 실행 결과와 검정통계량 계산식에 의한 이론 결과는 정확히 일치하였다.

3-2. 반복없는 이원분산분석법

하나의 요인 변수는 4개 그룹 (그룹당 3개 측정치), 다른 하나의 요인 변수는 3개 그룹 (그룹당 4개 측정치)으로 자료가 구성되었다.

<표 8> 반복없는 이원배치법 엑셀메뉴 실행결과

분산 분석: 반복 없는 이원 배치법						
요약표	관측수	합	평균	분산		
Row 1	3	88052.3	29350.77	37450726		
Row 2	3	82112.4	27370.8	66791410		
Row 3	3	48423.2	16141.07	20389830		
Row 4	3	68354.1	22784.7	29508437		
Column 1	4	115595	28898.75	48012360		
Column 2	4	102422.8	25605.7	39466439		
Column 3	4	68924.2	17231.05	21887976		
분산 분석						
변동의 요인	제곱합	자유도	제곱평균	F비	P-값	F기각치
인자 A(행)	3.1E+08	3	1.03E+08	32.94503	0.000401	4.757083
인자 B(열)	2.89E+08	2	1.45E+08	46.20617	0.000227	5.143253
잔차	18795253	6	3132542			
계	6.18E+08	11				

<표 9> 반복없는 이원분산분석 계산에 의한 이론 결과

Source	SS	df	MS	F-ratio	p-value
호선	309605071	3	103201690	32.94503	0.000401
요일	289485552	2	144742776	46.20617	0.000227
오차	18795253	6	3132542.1		
11					

표 8 과 표 9 에서 보듯이 반복없는 자료에 관한 2원분산 분석법은 엑셀 메뉴를 사용한 경우와 검정통계량 계산식에 의한 이론 결과가 일치하였다.

3-3. 반복있는 이원분산분석법

하나의 요인은 3개의 그룹을, 다른 하나의 요인은 4개의 그룹을 가진 변수이며, 4번 반복하여 자료가 측정되었다. 따라서 총 자료갯수는 48개 이다.

<표 10> 반복있는 이원분산배치 엑셀 실행 결과

분산 분석: 반복 있는 이원 배치법

요약표	평일	토요일	일요일	계
1호선				
관측수	4	4	4	12
평균	32220.325	32157.475	22328.05	28901.95
2호선				
관측수	4	4	4	12
평균	34465.33	28845.18	18781.075	27363.86167
3호선				
관측수	4	4	4	12
평균	19453.5	17608.50833	11125.30833	16062.43889
4호선				
관측수	4	4	4	12
평균	26567.17308	24634.08654	16680.04808	22627.10256
계				
관측수	16	16	16	
평균	28176.58202	25811.31247	17228.62035	

분산 분석

변동의 요인	제공합	자유도	제공 평균	F 비	P-값	F 기각치
인자 A(행)	1199538973	3	399846324.4	227.6312589	1.87025E-23	2.866265557
인자 B(열)	1061946499	2	530973249.7	302.2814063	3.12848E-23	3.259446306
교호작용	75253828.67	6	12542304.78	7.140294787	4.56138E-05	2.363750958
잔차	63235900.69	36	1756552.797			
계	2399975202	47				

<표 11> 반복있는 이원분산분석 계산식에 의한 이론 결과

Source	SS	df	MS	F	pvalue
호선	1199538	3	399846	227.6	1.87025E-23
요일	1061946	2	530973	302.2	3.12848E-23
교호	7525387	6	125423	7.140	4.56138E-05
오차	6323597	36	175655		

표 10 와 표 11 에서 보듯이 반복있는 이원분산분석법 경우에도 엑셀메뉴를 실행한 결과와 검정통계량 계산식에 의한 이론 결과는 정확히 일치한다.

3-4. 매크로 실행 결과

3-4-1. 블록화된 분산분석 결과

일원분산분석에서 사용하였던 자료에서 각 그룹내 자료들 순서를 블록으로 계산한 분산분석 결과이다.

<표12> 블록화된 분산분석 매크로 실행 결과

Source	SS	df	MS	V.R.	pvalue
Trt	1.04484	3	0.348	64.937	1.09097E-07
Blk	0.99132	4	0.247		
Error	0.06436	12	0.005		

앞의 표 6 과 표 7 의 일원분산분석 결과와 비교하면, 그룹간 총 자승합은 동일하다.

그러나, 일원 분산분석에서의 그룹내(오차) 총 자승합이 블록 자승합과 오차 자승합으로 분리되어 오차 자승합이 감소된다. 따라서 요인의 검정통계량 값이 증가하고 p 값은 감소하여 요인 효과를 설명할 가능성이 높아진다.

3-4-2. 일원분산분석의 다중비교 매크로 실행 결과

<표 13> 그룹간 평균 비교 결과

	A	B	C	D
25	Multiple Comparison			
26				pvalue
27	1 vs 2	-2.13311	2.133105457	0.04875
28	1 vs 3	2.029191	2.029191339	0.059415
29	1 vs 4	-1.20007	1.200072462	0.247579
30	2 vs 3	3.580939	3.580939334	0.002498
31	2 vs 4	0.937449	0.937449142	0.362464
32	3 vs 4	-0.30304	0.30303748	0.765766
33				

표 6 과 표 7 의 일원분산분석 결과에서 모든 그룹의 평균이 동일하다는 영가설이 기각되었다 (p=0.01). 따라서 어느 평균의 차이가 있는 그룹들 쌍을 찾아내는 과정이 중요하다. 실행 결과에서 1그룹과 2그룹, 2그룹과 3그룹 평균의 차이가 있음이 검정되었다.

4. 결론

두 개 이상 그룹의 평균값을 비교하는 분산분석법은 자료분석에서 기본적으로 많이 사용되는 통계분석 방법이다.

엑셀 프로그램에 존재하는 자료분석 메뉴 활용도를 높이기 위하여 본 연구에서는 결과의 정확성을 분석하고, 또한 분산분석에 많이 사용되는 분석방법 결과들을 얻기 위한 매크로를 개발하였다.

(1) 두 개 이상의 그룹을 가진 요인이 한 개인 일원분산분석 자료를 엑셀 일원배치 메뉴 실행 결과와 일원분산분석법 검정통계량에 의한 이론 결과는 정확히 일치하였다.

즉 일원배치 메뉴의 정확성이 검정되었다.

(2) 반복 측정 되지 않은 요인이 두 개인 이원분산분석 자료에서도 엑셀 메뉴 실행 결과와 이원분산분석법 검정통계량에 의한 이론 결과는 정확히 일치하였다.

즉 반복없는 이원배치법 메뉴의 정확성이 검증되었다.

(3) 반복측정된 요인이 두 개인 이원분산분석 자료에서도 엑셀메뉴 실행 결과와 이원분산분석법 검정통계량에 의한 이론 결과는 정확히 일치하였다.

즉 반복된 이원배치 메뉴의 정확성이 검증되었다.

(4) 일원분산분석법에서 오차를 줄이기 위하여 블록이란 요소를 적용한 분석법은 아직 엑셀 메뉴에는 존재하지 않는다.

그러나 본 연구에서는 엑셀의 함수를 이용하여 블록화된 분산분석 결과를 얻을 수 있는 매크로를 개발하였다.

일원분산분석에서 사용되었던 자료를 개발된 매크로에 실행한 결과 오차가 감소한 결과를 얻을 수 있었다.

(5) 모든 그룹의 평균이 동일하다는 가설이 기각될 때, 평균의 차이가 있는 그룹을 찾기위한다중 비교 결과를 얻기 위한 매크로를 개발하였다.

일원분산분석법에서 사용되었던 실 데이터에 적용 시킨 결과 두 쌍 그룹들에서 차이가 발견되었다.

통계학-지식월드	2008.11.25
[6]	
http://cafe.naver.com/anova99.cafe?iframe_url=/ArticleRead.nhn%3Farticleid=138	
반복측정자료의분산분석법	2008.11.05

김숙영

미국 오하이오 주립대학 컴퓨터 과학과 졸업
미국 오하이오 주립 대학원 석사 (응용통계학)
현재 안산공과대학 컴퓨터정보과 재직

관심분야: 전산통계, 전산회계, 전산수학

참 고 문 헌

[1]
http://cafe.naver.com/kgmath.cafe?iframe_url=/ArticleRead.nhn%3Farticleid=50
통계학 역사 2008.5.31.

[2]
<http://www.refee.com/search/?rq=11&qt=%C5%EB%B0%E8%C7%D0%B0%B3%B7%D0>
통계학 개론 2008.11.18

[3]
http://kin.naver.com/detail/detail.php?dlid=1&ir_id=1050202&eid=8IgHCWZFsanzklFN/x/amttNMSz0cPLVV&qb=v6K8v7WlwMzFzbrQvK6x4rTJ&pid=fR822woi5TCssb/pxtoss--480659&sid=SU82WCctT0kAACakqxM 엑셀2007의 통계분석기능, 2007.11.15

[4]
<http://blog.naver.com/jooning?Redirect=Log&logNo=40056022178> 엑셀 2007 매크로, 2008.10.16

[5]
<http://www.refee.com/search/?rq=11&qt=%C5%EB%B0%E8%C7%D0>