CHAPTER 9

분산 분석

9.1. 분산 분석 개념

분산분석(ANOVA: Analysis of Variance)이란 종속변수(dependent variable: 반응변수: response variable)의 분산(variation: 변동 ➔ 통계에서는 이를 변수가 가진 정보라 한다)을 설명하는 독립 변수 (independent: 설명변수: explanatory)의 유의성(significant: 설명력이 있는가)을 알아보는 방법이다. 반응 변수는 측정형 변수이고 독립 변수들은 분류형 변수이어야 하며 이를 처리 효과(treatment effect)혹은 인자(factor)라 한다.

만약 설명 변수가 측정형이면 회귀분석(10 장)을 실시하고 종속 변수가 분류형이면 교차 분석(8 장)을 실시하면 된다.

설명 변수의 유의성은 F-검정을 하고 각 설명 변수의 수준 간 차이는 Ad-hoc(사후 검정) 분석 혹은 다중 비교(multiple comparison)를 실시한다. 다중 비교 방법은 Tukey 방법, Scheffe 방법, Duncan 방법 등이 있는데 사회 과학에서는 Scheffe 방법, 자연 과학에서는 Tukey 방법을 가장 많이 사용한다. Tukey 방법이 가장 보수적인 방법(귀무가설을 기각할 가능성이 낮다)이기 때문이다.

예를 들어 바이러스 수(개수: 종속 변수, 측정형)에 영향을 미치는 요인으로 온도(10 도, 20 도, 30 도: 분류형)와 습도(10%, 30%, 50%: 분류형)를 고려하여 자료를 수집하였다고 하자. 물론 자료를 수집하는 (온도, 습도)의 수준 결합에 따라 바이러스 수를 측정하면 된다. (실험계획법) 이 예제는 2 원 분산분석(two-way ANOVA: 요인이 2 개)이 된다. 우선 분산 분석표(F-검정)에 의해 요인 둘 중 적어도 하나가 유의한지를 검정하고 Type III 에 의해 각 요인의

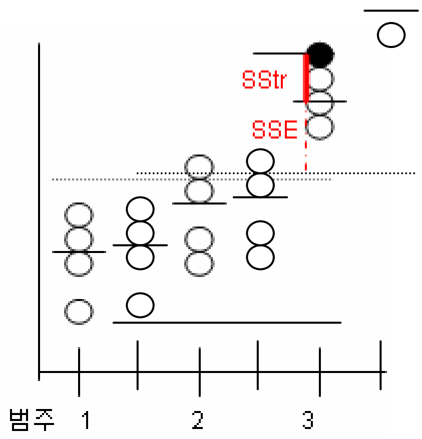
유의성을 검정한다. 온도 10 도, 20 도, 30 도 간 차이는 있는가, 그리고 습도 10%, 30%, 50% 간 바이러스의 차이는 있는가를 각각 알아보는 방법은 다중 비교 혹은 사후 검정이라 한다.

(1)모형 $y_{ij} = u + \tau_i + e_{ij} = \mu_i + e_{ij}$, $i = 1,2,\ldots k$ $j = 1,2,\ldots,n_i$, (i=요인의 수준 수, j=반복 수)

- y_{ij} = i 범주(집단)의 j 번째 관측치,
- μ = 모집단 총 평균 ➔ $\bar{y}$ (표본 자료 전체 평균)로 추정
- μ_i = i 범주의 평균 (i 집단 평균에 의해 설명되는 부분): $\bar{y}_i$ 로 추정(요인 수준 i 표본 자료 평균)
- e_{ij} = 집단 평균에 의해 설명되지 않는 부분

(2)변동: $\sum\sum(y_{ij} - \bar{y})^2 = \sum\sum(y_{ij} - \bar{y}_i)^2 + \sum\sum(\bar{y}_i - \bar{y})^2$

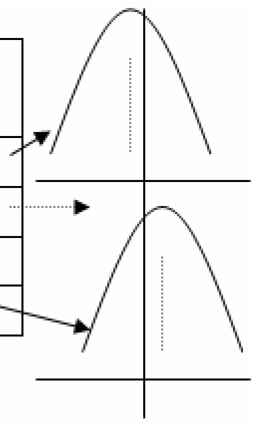
$SST = SSE + SStr$ (⇔ 총 변동=설명 안 되는 변동+설명 되는 변동)



9.1.1. Layout

서로 독립인 모집단이 3 개 이상인 경우 중앙 위치를 비교 분석은 이는 일원 배치(one-way layout) 실험 계획에 대한 분석 방법과 같다. 요인의 수준이 k 인 일원 배치 실험과 k 개의 모집단 중앙 위치에 대한 비교를 위한 자료 수집을 비교하면 다음과 같다.

	일원 배치 설계				모집단 중앙 위치 비교
요인	1	2	…	반복 수	
1	y_{11}	y_{12}	$\cdots y_{1n1}$	n_1	y_{11} y_{12} $\cdots$ y_{1n1}
2	y_{21}	y_{22}	$\cdots y_{2n2}$	n_1	y_{21} y_{22} $\cdots$ y_{2n2}
…	…	…		…	…
k	y_{k1}	y_{k2}	$\cdots y_{knk}$	n_k	y_{k1} y_{k2} $\cdots$ y_{1nk}



총 평균 (grand mean)

요인(처리 효과)과 반응 변수의 관계를 본다는 것은 요인 수준에 따른 반응 변수 평균 값의 차이가 유의 한가를 검정 하는 것과 같다.

9.1.2. 실험 설계

요인의 수준이 t 개이고 각 수준의 반복 수를 n_t 라고 한다면 다음과 같은 실험 설계 표를 만들고 각 셀에 1-n 까지 번호를 차례로 매긴다. 난수표에 의해 실험 순서를 정하고 실험을 실험 단위(unit)에 실시하면 된다. 그러므로 일원 분산의 실험 설계는 각 셀에 번호를 부여 한 후 난수를 이용하여 실험 순서를 결정하므로 이를 완전 임의 실험 설계(CRD: Completely Randomized Design)라 부르기도 한다.

비료 종류(A, B, C)에 따른 벼 수확량의 차이를 알아보기 위하여 실험을 한다고 하자. 반복 수는 각 비료에 대해 4 번을 한다면 실험을 위해 총 12 곳의 경작지가 필요하다.

①	②	③
④	⑤	⑥
⑦	⑧	⑨
⑩	⑪	⑫

12 경작지가 땅의 비옥도가 동일하다면 CRD 방법을 이용하여 실험하면 된다. 난수표를 이용하여 임의로 비료를 배정할 수도 있지만 총 실험 수가 많지 않으므로 12 장의 종이에 각

번호를 적고 던진 후 하나씩 선택하여 실험을 배정하면 된다. 만약 1, 5, 2, 7, 12, 10 …, 3 순으로 선택하였다면

① A	② A	③ C
④ B	⑤ A	⑥ B
⑦ A	⑧ C	⑨ C
⑩ B	⑪ C	⑫ B

9.1.3. 일원 분산분석표 (One-way ANOVA Table)

변동 source	자유도 DF	자승합 Sum of Square	평균 자승합 Mean of SS	F-검정 F-Stattistic
처리 효과 Treatment	$k-1$	$SStr=\sum\sum(\bar{y}_i-\bar{y})^2$	$MStr=SStr/(k-1)$	$F=\frac{MStr}{MSE}$ ~ $F(t-1,n-k)$
오차 Error	$n-k$	$SSE=SST-SStr$	$MSE=SSE/(n-k)$	
총합 Total	$n-1$	$SST=\sum\sum(y_{ij}-\bar{y})^2$		

그러므로 F 검정 통계량의 값이 커지면 처리 효과가 존재한다고 할 수 있다. 처리 효과가 존재한다는 것은 각 수준별 반응 변수 평균의 차이가 유의 하다는 것이다. F-검정은 귀무가설 $H_0:\mu_1=\mu_2=...=\mu_k$ 의 유의성을 검정하므로 귀무가설이 기각 된다는 의미는 집단간 차이가 있는 것을 의미하지 어떤 집단간 차이가 존재하는지 알 수 없다. 집단간 차이를 보기 위하여 사후 검정(Post HOC.) 혹은 다중 비교를 실시하게 된다.

만약 범주가 2 개인 경우는 (예를 들면 우리 설문 예제에서 성별에 따른 시설물 만족도 차이가 있는가?) 독립인 두 집단 평균 비교를 위한 모평균 차이 t-검정과 분산 분석은 동일하다. ($t^2(df=n)=F(df=(1,n))$)

9.1.4. 사후 검정 (Post-Hoc test) 혹은 다중 비교 (multiple comparison)

분산 분석의 F-검정은 단지 귀무가설 $H_0:u_1=u_2=...=u_k$ 즉 전체적인 차이를 검정하는 것이다. 그러므로 수준별 차이(pair wise: 예: $H_0:u_1=u_3$)가 있는지 혹은 수준의 선형 결합

대비(contrast: 예: $H_0 : u_1 = \frac{u_2 + u_3}{2}$)의 차이가 있는지 검정할 필요가 있는데 이를 사후 검정 혹은 다중 비교(대비 포함)라 한다. 사후 검정이므로 F-검정 결과와 관계없이 (귀무가설을 채택하더라도) 시행하게 된다. 다중 비교에서는 여러 개의 가설을 동시에 검정하므로 유의 수준을 조정해야 한다. 이를 조정된 실험 유의수준 (controlled experimental error rate)이라 하고 $1-(1-\alpha)^c$ 이다. 여기서 c 는 가설 수를 의미하며, pair wise 다중 비교의 경우 $c = k(k-1)/2$ 가 된다.

(1)Fisher's Least Significant Difference

pair wise (두 수준별 평균 비교) 검정에 사용한다. LSD 를 구하고 평균의 차이가 그보다 크면 귀무가설을 기각한다. 이는 다중 비교 방법에 해당되지 않는다.

(2)Tukey W procedure: studentized range distribution

$W = \max(y_i) - \min(y_i)$ 이용하여 $q = w/s$ 표 작성한다. 가장 보수적인 방법으로 자연 과학에서 가장 많이 이용한다.

(3)Student-Newman-Keuls procedure

Tukey 방법의 변형한 것으로 표만 다르다. 즉 임계치(critical value)만 차이가 있고 방법은 Tukey 와 동일하다.

(4)Duncan Multiple range test

Tukey 방법과 매우 유사하나 수준별 표본 평균을 크기 순으로 나열하여 차이가 가장 큰 것을 비교해 가면서 유의 수준을 $1-(1-\alpha)^r$ 으로 조정해 가면서 검정한다. r 은 검정 단계 순서이다. 귀무가설을 기각할 확률이 매우 높아 자주 사용하지 않는다.

(5)Scheffe's S method

대비(contrast)까지 고려한 다중 비교 방법이다. 사회 과학 분야 가장 많이 이용된다.

(6)Dunnett's procedure

처리 효과의 수준 하나가 control (실험 집단)인 경우 (예: placebo 집단, 교육을 하지 않는 집단, 이전 약 투여 집단) 이 집단과 다른 집단들을 pairwise 비교할 경우 사용된다.

9.1.5 예제 자료

다음은 3 개 호수의 산소량의 차이가 있는지 알아보기 위하여 각 호수의 중앙에서 깊이 1m 의 물로부터 산소량(ppm)을 측정한 자료이다. 호수에서 위치에 따라 산소량의 차이가 있을 것이므로 10 곳을 선택하여 각 산소량을 측정한 것이다.

Lake	Observation
1	0 2 1 3 1 2 3 4 1 5
2	1 3 4 6 8 7 5 3 4 5
3	14 26 25 18 19 22 21 16 20 30

⏭등분산 가정 검정: 분산분석에서 등분산 가정은 Hartley's test 를 이용하면 된다. 그러나 일반적으로 분산 분석에서는 등분산 가정을 검정하지 않는다.

▪귀무가설: $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$

▪검정통계량: $F_{\max} = \dfrac{\max(s_i^2)}{\min(s_i^2)} \sim F(\text{자유도}: \text{분자표본크기}-1, \text{분모표본크기}-1)$

```
data one;
     input lake oxygen @@;
     cards;
1 0 1 2 1 1 1 3 1 1 1 2 1 3 1 4 1 1 1 5
2 1 2 3 2 4 2 6 2 8 2 7 2 5 2 3 2 4 2 5
3 14 3 26 3 25 3 18 3 19 3 22 3 21 3 16 3 20 3 30
run;

proc glm data=one;
     class lake;
     model oxygen=lake;
     means lake/scheffe lines;
run;
```

means 문장은 다중 비교를 위한 것이다. LINES 옵션은 집단간 평균의 차이가 유의한 것을 구별하기 위해 사용되었다.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2117.400000	1058.700000	105.52	<.0001
Error	27	270.900000	10.033333		
Corrected Total	29	2388.300000			

변동	자유도	자승합 SS	평균자승합 MS	F-값	유의 확률 p-값
처리 효과	2	2117.4	1058.7	105.52	<.0001
오차	27	270.9	10.033333		
총합	29	2388.3			

유의 확률이 0.0001 미만이므로 호수에 따른 산소량의 차이는 매우 유의하다. 이제 호수 별 차이를 보기 위하여 Scheffe 다중 비교 결과를 보자. 다중 비교는 분산 분석표의 F-검정 결과가 유의하지 않더라도 실시한다.

Means with the same letter are not significantly different.

Scheffe Grouping	Mean	N	lake
A	21.100	10	3
B	4.600	10	2
B			
B	2.200	10	1

알파벳이 같은 범주(집단은) 차이가 없음을 의미한다. 호수 3 의 산소량 평균은 21, 호수 2 는 4.6, 호수 1 은 2.2 이다. 호수 1 과 2 의 산소량 차이는 없으나(알파벳이 A 로 동일) 호수 3 과 호수 2, 호수 3 과 호수 1 은 산소량의 차이가 존재한다. 집단 별(호수) 산소량의 평균과 표준 편차를 구하려면 TABULATE procedure 를 사용하는 것이 편리하다.

```
PROC TABULATE DATA=ONE FORMAT=5.2;
    CLASS LAKE;
    VAR OXYGEN;
    TABLE (LAKE ALL),OXYGEN*(MEAN STD);
RUN;
```

	oxygen	
	Mean	Std
lake		
1	2.20	1.55
2	4.60	2.07
3	21.10	4.84
All	9.30	9.07

9.2. 설문 분석에 적용

설문 조사에서는 분산분석이 이용되는 경우는 리커드 척도 문항의 점수가(종속 변수) 인구학적 변인, 혹은 일반 객관식 선택 문항(설명 변수, 요인)에 따라 차이가 있는지 알아보고자 할 때이다. 일반적으로 분산 분석은 요인(설문에서는 문항으로 생각하면 된다)이 2 개 이상인 경우도 다룬다. 이를 2 원(two-way) 분산 분석, 다원(multi-way) 분산 분석이라 한다. 설문 분석에서도 인구학적 문항이 여러 개인 경우가 많아 이 문항을 동시에 고려하는 다원 분산 분석이 필요할 수 있다. 그러나 정말 필요한가? ①교차 항(두 요인간 결합)을 고려하는 경우 해석이 용이하지 않고 ②실험 계획에 의해 얻어진 자료가 아니므로 각 문항을 하나씩 개별적으로 분석하는 일원 분산 분석(요인이 하나)이면 충분하다. 그러므로 설문 조사에서 분산 분석을 사용할 때는 분류형 문항(일반적으로 인구학적 변인 문항)하나, 리커드 문항인 종속 변수인 일원 분산 분석을 실시한다. 설문 조사 분석에서 분산 분석을 실시한다는 것은 인구학적 변인에 따른 리커드 척도 문항의 평균 점수 차이가 있는지 결정하는 것이다. 예를 들어 선택하는 전공(Q27: 중국, 경제, 정보 통계)에 따른 입학한 것에 따른 만족도(Q24)의 차이는 있는지를 알아보고자 할 때 분산 분석을 실시하면 된다. 이 때 분산 분석을 실시한다는 것은 중국 전공 선택 학생들의 입학 만족도 평균, 경제 선택 전공자의 만족도 평균, 정보통계 전공자의 만족도 평균의 차이가 있는지 알아 보는 것이다.

예제 설문에서 다음을 실시해 보자.

- 성별(Q1)에 따른 ○○대학교 전체적 만족도(Q25)의 차이가 있는가? (F 검정: 특히 수준이 2 개인 경우 F-검정은 t-검정과 동일하다. 왜냐하면 $F(df=1,n)=t^2(n)$ 이므로, 즉 두 모집단 평균 차이 비교 t-검정은 F-검정과 동일하다.) 여기서는 집단의 범주가 2 개인 경우에는 PROC TTEST 를 이용하여 분석하기로 한다.
- 선택한 전공(Q27)에 따른 ○○대학교 전체적 만족도(Q25)의 차이가 있는가?
- 선택한 전공(Q27)에 따른 요인 분석 결과 묶은 그룹 만족도(강의실 만족도: LECTURE Q5, Q6, Q7, Q8)의 차이는 있는가?

9.2.1. SAS

(1)기초 통계량

우선 문항의 범주 각각에 대한 평균과 분산을 구하면 보고서 작성에 유리하다.

```
DATA SURVEYO;
    SET SURVEY;
        LECTURE=MEAN(OF Q5-Q8);
        INFORMATION=MEAN(Q10, Q11);
RUN;
```

강의실 만족도(LECTURE)와 정보 시설 만족도를 구하였다.

```
PROC TABULATE DATA=SURVEYO;
    CLASS Q1;
    VAR Q25;
    TABLE Q1 ALL, Q25*(MEAN STD);
RUN;
```

	Q25	
	Mean	Std
Q1		
1	3.79	1.28
2	4.29	1.16
All	3.94	1.26

Q1 의 범주(class)에 의해 변수(문항) Q25 의 평균과 분산을 (mean, std) 출력하게 한다. ALL 의 의미는 범주를 고려하지 않고 총 평균을 구하는 것이다. TABLE 문에서, 앞은 행을 뒤는 열을 의미하며 *의 의미는 교차의 의미이다. 여자(Q1=1)에 비해 남자의 한남대 만족도(Q25)가 높다. 물론 유의적 차이 검정은 t-검정(분산 분석) 해야겠지만…

```
PROC TABULATE DATA=SURVEYO;
    CLASS Q27;
    VAR LECTURE Q25;
    TABLE Q27 ALL, (LECTURE Q25)*(MEAN STD);
RUN;
```

	LECTURE		Q25	
	Mean	Std	Mean	Std
Q27				
1	2.75	0.91	3.83	1.25
2	2.58	1.46	4.33	1.41
3	2.73	1.01	4.40	0.97
All	2.74	0.95	3.91	1.25

강의실 만족도는 중국 전공(Q27=1)이 ○○대 만족도(Q25)는 정보 통계 전공이 가장 높았다. 물론 통계적 유의적 차이에 대한 검정은 분산 분석을 해야 하지만…

(2)t-검정(집단이 두 개)

집단이 2 개인 경우에는 t-검정이 실시하면 된다. 다음은 Q1(성별)에 의한 한남대학교 만족도 차이가 있는 t-검정을 실시한 프로그램이다.

```
PROC TTEST DATA=SURVEY;
    CLASS Q1;
    VAR Q25;
RUN;
```

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
Q25	Pooled	Equal	125	-2.08	0.0396
Q25	Satterthwaite	Unequal	76.9	-2.17	0.0334

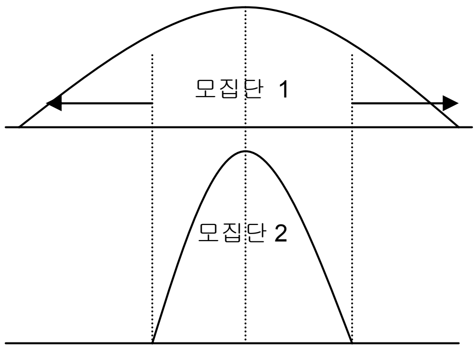
Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
Q25	Folded F	88	37	1.22	0.4988

t-검정에서는 두 모집단의 분산이 같은지 검정하여 만약 등분산 가정이 만족하면 일반적인 t-검정 방법을 등분산 가장이 만족하지 않으면(이분산) 수정된 t-검정(satterthwaite)을 실시한다. 위의 예제에서 등분산 가정 유의성 검정 결과 유의 확률 p-값이 0.4988 이므로 Method=Pooled 방법의 검정(일반적인 t-검정)을 사용하여 두 모집단의 평균 차이는 있다고 판별한다.(p-값=0.0396) 그러므로 남자가 여자에 비해 한남대학교에 대한 만족 정도가 높다. (이전 페이지: 남자=4.29, 여자는 3.79)

※두 모분산이 같다는 가정이 왜 필요한가?

모집단1과 모집단2의 평균은 동일하지만 모집단1로부터 뽑은 표본은 화살표 부분에서도 관측된다. 그러므로 표본에 의한 가설 검정 결과 모집단 평균은 같지 않다는 결론에 도달 할 수 있다. 그러므로 모분산이 같은 경우와 같지 않은 경우 나누어 검정하게 된다. 등분산 가정 검정은 두 집단의 표본 분산 비를 이용하여 검정한다.



(3)분산 분석(ANOVA: 3 집단 이상)

```
DATA SURVEY0;
    SET SURVEY;
        LECTURE=MEAN(OF Q5-Q8);
        INFORMATION=MEAN(Q10, Q11);
RUN;

PROC GLM DATA=SURVEY0;
    CLASS Q27;
    MODEL Q25 LECTURE=Q27;
    MEANS Q27/LINES SCHEFFE;
RUN;
```

CLASS 문은 요인 변수 지정한다.

MODEL Q25=Q27과 MODEL LECTURE =Q27 모형 2개를 분석한다.

Dependent Variable: Q25

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	4.6271646	2.3135823	1.50	0.2282
Error	125	193.4275229	1.5474202		
Corrected Total	127	198.0546875			

선택한 전공에 따른 한남대학교 만족도의 차이는 없다.

Scheffe Grouping	Mean	N	Q27
A	4.4000	10	3
A			
A	4.3333	9	2
A			
A	3.8349	109	1

Scheffe 사후 분석 결과 각 집단간 pairwise 차이도 존재하지 않았다. 출력 결과가 Q27 문항의 보기 순이 아니라 평균 크기 순임을 유의하기 바란다. Q27=3(정보통계)의 한남대 만족도 4.4 로 가장 높았고 Q27=1(중국전공) 3.83 으로 가장 낮았다. 물론 통계적인 유의서 차이는 없었다. 선택한 전공에 따른 강의실 만족도 차이도 없었으며 Scheffe 사후 분석 결과 집단간 pairwise 차이도 존재하지 않았다. (결과 생략)

만약 사후 검정 결과 차이가 있다면 다음과 같이 나타난다.

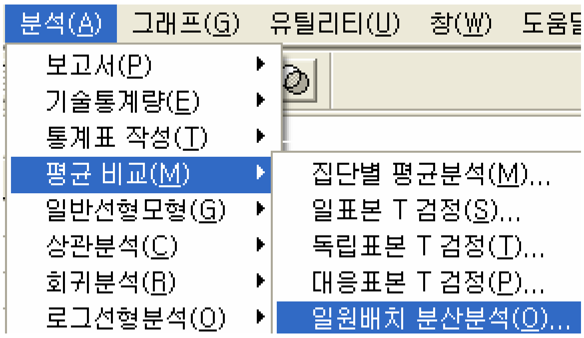
Scheffe's Test for life

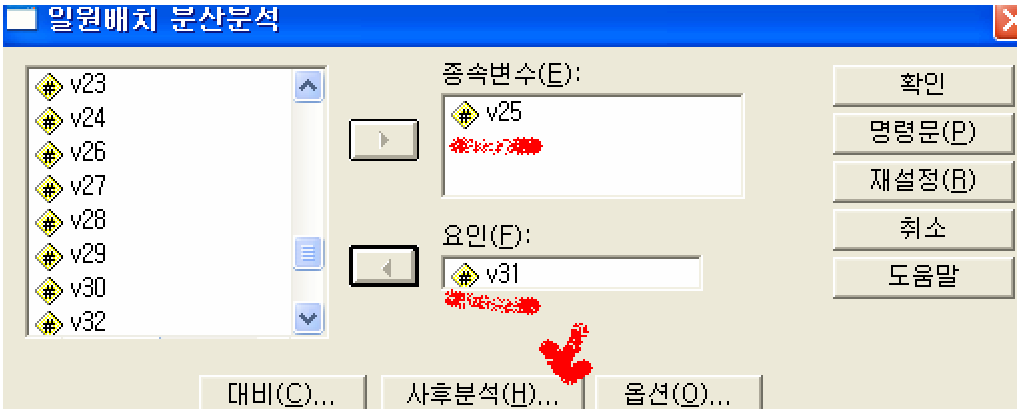
ng		Mean	N	age
	A	1.3037	37	60+
B	A	1.0557	31	80+
B		1.0071	45	70+

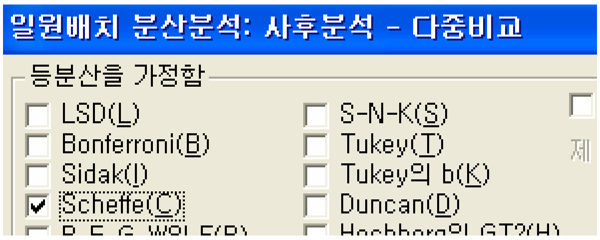
변수 age 가 60 대(60+)과 80 대, 80 대와 70 대의 차이는 유의하지 않으나 60 대와 70 대는 차이가 있음을 알 수 있다.

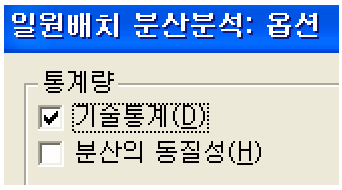
9.2.2. SPSS

전공 선택에(V31) 따른 한남대학교 만족도 V25 의 차이가 있는지 분산 분석하자.









기술통계

V25

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰구간 하한값	평균에 대한 95% 신뢰구간 상한값	최소값	최대값
1	109	3.83	1.25	.12	3.60	4.07	1	7
2	9	4.33	1.41	.47	3.25	5.42	3	7
3	10	4.40	.97	.31	3.71	5.09	3	6
합계	128	3.91	1.25	.11	3.70	4.13	1	7

분산분석

V25

	제곱합	자유도	평균제곱	F	유의확률
집단-간	4.627	2	2.314	1.495	.228
집단-내	193.428	125	1.547		
합계	198.055	127			

다중 비교

종속변수: V25
Scheffe

					95% 신뢰구간	
(I) V31	(J) V31	평균차 (I-J)	표준오차	유의확률	하한값	상한값
1	2	-.50	.43	.515	-1.57	.57
	3	-.57	.41	.391	-1.58	.45
2	1	.50	.43	.515	-.57	1.57
	3	-6.67E-02	.57	.993	-1.48	1.35
3	1	.57	.41	.391	-.45	1.58
	2	6.67E-02	.57	.993	-1.35	1.48

각 집단간 차이는 유의하지 않았다. 모든 유의 확률 값(p-값)이 0.05 보다 많이 크다.

9.3. 보고서 작성

보고서 작성은 다음과 같다. 만약 출신 지역(서울, 경기, 그 외 지역을 하나로 묶어 3 지역)에 따른 한남대학교 수준(Q22) 만족도의 차이가 있는지 알아보자.

```
DATA SURVEY0;
    SET SURVEY;
    IF (Q3=>3) THEN Q3=3;
RUN;
```

▶서울, 경기, 그 외 지역을 하나로 묶는 프로그램.

```
PROC GLM DATA=SURVEY0;
    CLASS Q3;
    MODEL Q22-Q25=Q3;
    MEANS Q3/LINES SCHEFFE;
RUN;
```

▶MODEL Q22=Q3, MODEL Q23=Q3, 4개 모형에 대한 분산 분석을 실시한다.

다음은 4 개 분산 분석 중 MODEL Q22=Q3 결과이다.

Dependent Variable: Q22

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	12.8423621	6.4211810	5.05	0.0077
Error	126	160.0568627	1.2702926		
Corrected Total	128	172.8992248			

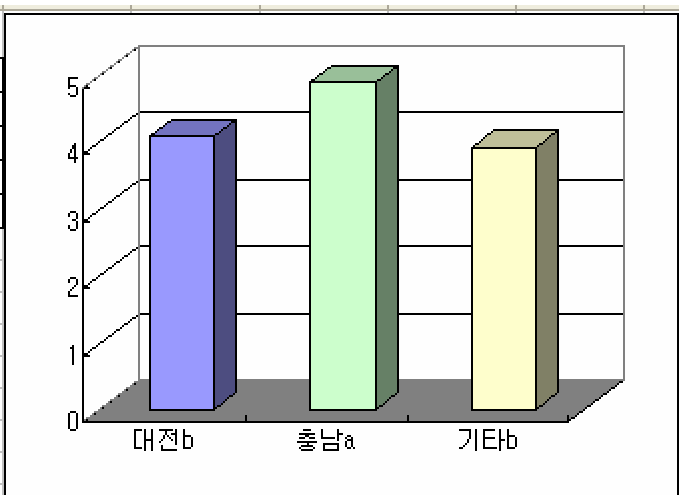
Means with the same letter are not significantly different.			
Scheffe Grouping	Mean	N	Q3
A	4.9000	20	2
B	4.0824	85	1
B	3.9167	24	3

보고서를 작성할 때는 평균은 물론 분산도 필요하므로 다음 프로그램을 돌린 후 분산 분석 결과를 이용하여 표를 완성한다.

```
PROC TABULATE DATA=SURVEY0;
    CLASS Q3;
    VAR Q22;
    TABLE Q3 ALL, Q22*(MEAN STD);
RUN;
```

	Q22	
	Mean	Std
Q3		
1	4.08	1.13
2	4.90	0.91
3	3.92	1.28
All	4.18	1.16

한남대 수준 만족도		
출신지역	Mean	Std
대전b	4.08	1.13
충남a	4.9	0.91
기타b	3.92	1.28
All	4.18	1.16



(F – 통계량 : 5.05, 유의확률 : 0.0077

지역 옆의 알파벳은 Scheffe 의 다중 비교 결과를 표시한 것이다. 충남 지역의 출신자의 한남대 수준 만족도는 가장 높고 다른 지역과 차이가 유의함을 알 수 있다. 분산 분석의 F-통계량과 유의 확률(p-값)은 표나 그래프 옆이나 아래 적어주면 된다.

[연습문제]

(1)출신 지역(대전, 충남, 기타 지역 3 집단)에 따른 교양 영어(의사소통 영어: Q20)에 대한 만족도 차이가 있는지 분산 분석하시오. 사후 검정 방법으로 Scheffe 방법을 이용하시오.

(2)7 장 요인 분석 결과 교양 과목 만족도 4 개 문항 묶인 결과에 대해 출신 지역에 따른 차이가 있는지 분산 분석을 실시하시오. 사후 검정 방법으로 TUKEY 방법을 이용하시오.

(3)팀 프로젝트 설문지에서 인구학적 변인에 따른 리커드 척도 문항의 차이가 있는지 분석하고 결과를 정리하시오. 요인 분석 결과 묶인 문항은 그룹에 대해 분산 분석하시오.