The Korean Communications
in Statistics Vol. 7, No. 1, 2000
pp. 179-198

MINITAB Macros for Testing the Difference of Mean Vectors of Two Multivariate Populations

Hyuk Joo Kim[1] and Min Ah Ji[2]

Abstract

We consider the problem of comparing the mean vectors of two multivariate populations. We focus on testing hypotheses concerning two multivariate mean vectors by use of MINITAB. For the cases of small sample and large sample, MINITAB programs and outputs are presented for solving statistical problems. The MINITAB programs made in this paper are saved as macro files, and thus can be conveniently used for solving another problems.

1. 서 론

통계학의 중요한 논제이며 일상에서 흔히 접할 수 있는 문제 중의 하나는 두 집단을 비교하는 문제이다. 특히 두 모집단으로부터 표본을 뽑아 표본평균의 차를 구하여 모집단의 평균을 비교하는 문제는 통계적 추론의 대표적 문제 중 하나이며, 이미 개발된 방법들이 널리 사용되고 있다. 하나의 변수만을 다루는 일변량 문제의 경우에는 자료와 계산이 간단하기 때문에 별 문제가 없으나, 고려되는 변수가 많아져서 다변량 문제가 되면 자료가 방대해지고 계산도 복잡해져서 컴퓨터의 도움이 없이는 처리하기가 힘들어진다.

본 논문에서는 통계패키지 미니탭(MINITAB)을 이용하여 두 다변량모집단의 평균벡터의 차에 관한 가설을 검정하는 방법에 관한 문제를 다루고자 한다. 제2절에서는 전통적인 방법에 관한 일반적인 내용을 소표본의 경우와 대표본의 경우로 나누어 소개하고, 제3절에서는 두 가지 경우 각각에 대한 데이터를 예로 제시한 후 미니탭 프로그램을 작성하여 이 데이터를 처리하는 과정을 보일 것이다. 끝으로 제4절은 결론과 주목할 점으로 이루어진다.

2. 두 다변량모집단의 평균벡터의 차에 관한 검정

이 절에서는 두 다변량모집단의 평균벡터의 차에 관한 가설을 검정하는 방법에 관한 일반적인 내용을 설명한다. 이 방법은 Anderson(1984), Johnson과 Wichern(1982), 강병서(1988) 등의 일반

1) Associate Professor, Division of Mathematical Sciences, Wonkwang University, Iksan, Chonbuk, 570-749, Korea
2) Master, Major of Mathematics Education, Graduate School of Education, Wonkwang University, Iksan, Chonbuk, 570-749, Korea

적인 다변량통계학 서적에 소개되어 있다.

평균벡터가 $\underline{\mu}_1$, 공분산행렬이 Σ_1인 p변량모집단으로부터의 크기 n_1인 확률표본을 $\underline{X}_{11}$, $\underline{X}_{12}$, …… , $\underline{X}_{1n_1}$ 으로 나타내고, 평균벡터가 $\underline{\mu}_2$, 공분산행렬이 Σ_2인 p변량모집단으로부터의 크기 n_2인 확률표본을 $\underline{X}_{21}$, $\underline{X}_{22}$, ……, $\underline{X}_{2n_2}$ 로 나타내자. 여기서 각각의 $\underline{X}_{ij}$ $(i=1,2$; $j=1,2,\cdots\cdots, n_i)$는 p개의 변수들로 이루어진 $p\times 1$ 벡터이다.

이 표본들의 평균벡터와 공분산행렬은 각각 다음과 같다.

$$\underline{\bar{X}}_1=\frac{1}{n_1}\sum_{j=1}^{n_1}\underline{X}_{1j}\,,\qquad S_1=\frac{1}{n_1-1}\sum_{j=1}^{n_1}(\underline{X}_{1j}-\underline{\bar{X}}_1)(\underline{X}_{1j}-\underline{\bar{X}}_1)'$$

$$\underline{\bar{X}}_2=\frac{1}{n_2}\sum_{j=1}^{n_2}\underline{X}_{2j}\,,\qquad S_2=\frac{1}{n_2-1}\sum_{j=1}^{n_2}(\underline{X}_{2j}-\underline{\bar{X}}_2)(\underline{X}_{2j}-\underline{\bar{X}}_2)'$$

두 개의 확률표본들이 서로 독립적이라고 가정하고, 모집단 1의 평균벡터와 모집단 2의 평균벡터의 차에 관한 가설

$$H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0$$
$$H_1 : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\delta}_0$$

를 검정하려 한다. 여기서

$$\underline{\mu}_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{pmatrix},\quad \underline{\mu}_2 = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{pmatrix}$$

이다.

대표본인 경우에는 중심극한정리에 의하여 표본은 정규분포에 근사하게 되어 더 이상의 가정은 필요없다. 그러나 소표본의 경우에는 몇 가지의 가정이 더 필요하다. 두 다변량모집단의 평균벡터의 차에 관한 가설을 검정하기 위하여 소표본의 경우와 대표본의 경우로 나누어 논의한다.

2.1 소표본의 경우

소표본의 경우, 즉 n_1과 n_2가 작은 경우에는 두 모집단의 분포가 다변량정규분포이고 두 모집단의 공분산행렬이 같다는 가정이 필요하다.

동일한 공분산행렬을 Σ로 나타내기로 하자(즉 $\Sigma_1=\Sigma_2=\Sigma$). 그러면 $\sum_{j=1}^{n_1}(\underline{X}_{1j}-\underline{\bar{X}}_1)(\underline{X}_{1j}-\underline{\bar{X}}_1)'$은 $(\ n_1-1)\Sigma$의 추정량이며, $\sum_{j=1}^{n_2}(\underline{X}_{2j}-\underline{\bar{X}}_2)(\underline{X}_{2j}-\underline{\bar{X}}_2)'$은 $(\ n_2-1)\Sigma$의 추정량이다.

Σ를 추정하기 위해 두 표본의 정보를 결합하여 얻어지는 합동추정량은

$$S_{pooled} = \frac{\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)(X_{1j}-\bar{X}_1)' + \sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)(X_{2j}-\bar{X}_2)'}{n_1+n_2-2}$$
$$= \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2}$$

이다. 여기서 $\sum_{j=1}^{n_1}(X_{1j}-\bar{X}_1)(X_{1j}-\bar{X}_1)'$는 n_1-1의 자유도를 가지며, $\sum_{j=1}^{n_2}(X_{2j}-\bar{X}_2)(X_{2j}-\bar{X}_2)'$는 n_2-1의 자유도를 가지므로 S_{pooled}의 분모는 두 자유도를 합한 것이 된다.

귀무가설 $\mu_1 - \mu_2 = \delta_0$ 을 검정하기 위해 $\bar{X}_1-\bar{X}_2$의 분포를 살펴보자. $\bar{X}_1-\bar{X}_2$의 평균벡터는

$$E(\bar{X}_1-\bar{X}_2) = E(\bar{X}_1)-E(\bar{X}_2) = \mu_1-\mu_2$$

이다. 또한 $\bar{X}_1$과 $\bar{X}_2$는 독립적이므로 $Cov(\bar{X}_1, \bar{X}_2)=0$ 이다. 따라서 $\bar{X}_1-\bar{X}_2$ 의 공분산행렬은

$$Cov(\bar{X}_1-\bar{X}_2)=Cov(\bar{X}_1)+Cov(\bar{X}_2)=\frac{1}{n_1}\Sigma+\frac{1}{n_2}\Sigma=\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma$$

이다.

합동추정량 S_{pooled}는 Σ의 추정량이므로 $\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{pooled}$는 $Cov(\bar{X}_1-\bar{X}_2)$의 추정량이 된다. 따라서 $\bar{X}_1-\bar{X}_2$는 p변량정규분포 $N_p\left(\mu_1-\mu_2, \left(\frac{1}{n_1}+\frac{1}{n_2}\right)\Sigma\right)$ 를 따른다.

귀무가설

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

의 우도비검정(likelihood ratio test)은 검정통계량

$$T_0^2 = (\bar{X}_1-\bar{X}_2-\delta_0)'\left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_{pooled}\right]^{-1}(\bar{X}_1-\bar{X}_2-\delta_0)$$
$$= \frac{n_1 n_2}{n_1+n_2}(\bar{X}_1-\bar{X}_2-\delta_0)' \quad S_{pooled}^{-1} \ (\bar{X}_1-\bar{X}_2-\delta_0)$$

를 이용하여 다음과 같이 주어진다.

$T_0^2 > c$ 이면 유의수준 α에서 H_0를 기각.

단, 여기서 c 는 이표본 T_0^2 통계량의 귀무분포로부터 결정되는 기각값으로서 다음의 정리에 의하여 주어진다.

정리 2.1 $\underline{X}_{11}, \underline{X}_{12}, \cdots\cdots, \underline{X}_{1n_1}$ 이 $N_p(\mu_1, \Sigma)$로부터의 크기 n_1인 확률표본이고 $\underline{X}_{21}, \underline{X}_{22}, \cdots\cdots, \underline{X}_{2n_2}$ 가 $N_p(\mu_2, \Sigma)$로부터의 크기 n_2인 독립인 확률표본이면,

$$T^2=[\bar{\underline{X}}_1-\bar{\underline{X}}_2-(\mu_1-\mu_2)]' \left[\left(\frac{1}{n_1}+\frac{1}{n_2}\right) S_{pooled}\right]^{-1}[\bar{\underline{X}}_1-\bar{\underline{X}}_2-(\mu_1-\mu_2)]$$

은

$$\frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F(p, n_1+n_2-p-1)$$

의 분포를 따른다. 따라서

$$P\left[\{\bar{\underline{X}}_1-\bar{\underline{X}}_2-(\mu_1-\mu_2)\}' \left\{\left(\frac{1}{n_1}+\frac{1}{n_2}\right) S_{pooled}\right\}^{1}\{\bar{\underline{X}}_1-\bar{\underline{X}}_2-(\mu_1-\mu_2)\}\le c\right]=1-\alpha$$

이다. 단, 여기서

$$c=\frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F(p, n_1+n_2-p-1; \alpha)$$

이며, $F(p, n_1+n_2-p-1; \alpha)$는 자유도가 p와 n_1+n_2-p-1인 F분포의 상위 $100\alpha\%$ 백분위수이다.

증명. Johnson과 Wichern(1982, p.240) 참조.

2.2 대표본의 경우

표본의 크기 n_1과 n_2가 변수의 개수 p에 비해 아주 커서 n_1-p와 n_2-p가 큰 경우에는 정규모집단의 가정과 등분산($\Sigma_1=\Sigma_2$)의 가정이 더 이상 필요하지 않게 된다.

이 절에서는 공분산행렬이 같지 않은 임의의 두 p변량모집단의 평균벡터를 비교하는 방법을 논의하고자 한다. 이 방법을 사용할 수 있기 위해서 n_1-p와 n_2-p의 값이 얼마나 커야 하는가에 관해서는 엄격한 기준은 없으나, 일반적으로 n_1-p와 n_2-p가 20 이상이면 사용할 수 있다고 알려져 있다.

평균벡터가 μ_1, 공분산행렬이 Σ_1인 p변량모집단으로부터의 크기 n_1인 확률표본과 평균벡터가 μ_2, 공분산행렬이 Σ_2인 p변량모집단으로부터의 크기 n_2인 확률표본의 표본평균벡터를 각각 $\bar{\underline{X}}_1$과 $\bar{\underline{X}}_2$라 하자. 두 모집단의 공분산행렬 Σ_1과 Σ_2는 같지 않고 n_1-p와 n_2-p는 크다고 가정한다. 두 표본평균벡터의 차 $\bar{\underline{X}}_1-\bar{\underline{X}}_2$의 평균벡터는

$$E(\bar{\underline{X}}_1-\bar{\underline{X}}_2) = \mu_1-\mu_2$$

이고, 공분산행렬은

$$\mathrm{Cov}(\bar{X}_1-\bar{X}_2) = \mathrm{Cov}(\bar{X}_1)+\mathrm{Cov}(\bar{X}_2) = \frac{1}{n_1}\Sigma_1+\frac{1}{n_2}\Sigma_2$$

이며, 중심극한정리에 의하여 $\bar{X}_1-\bar{X}_2$의 분포는 근사적인 p변량정규분포, 즉 $N_p\left(\mu_1-\mu_2,\ \frac{1}{n_1}\Sigma_1+\frac{1}{n_2}\Sigma_2\right)$ 이다.

$\mu_1-\mu_2$에 관한 가설을 검정하는 데에 다음의 정리가 사용된다.

정리 2.2 위의 가정 하에서

$$[\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)]'\left(\frac{1}{n_1}\Sigma_1+\frac{1}{n_2}\Sigma_2\right)^{-1}[\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)]$$

는 자유도가 p인 카이제곱분포 즉 $\chi^2(p)$를 근사적으로 따른다. 또한 Σ_1과 Σ_2 대신 표본공분산행렬 S_1과 S_2를 각각 사용한

$$[\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)]'\left(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2\right)^{-1}[\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)]$$

도 역시 $\chi^2(p)$를 근사적으로 따른다.

증명. Johnson과 Wichern(1982, p.138과 p.245) 참조.

정리 2.2를 이용하면 귀무가설

$$H_0 : \mu_1 - \mu_2 = \delta_0$$

의 검정방법은 검정통계량

$$T_0^2=(\bar{X}_1-\bar{X}_2-\delta_0)'\left(\frac{1}{n_1}S_1+\frac{1}{n_2}S_2\right)^{-1}(\bar{X}_1-\bar{X}_2-\delta_0)$$

를 이용하여 다음과 같이 주어진다.

$T_0^2 > \chi^2(p;\alpha)$이면 유의수준 α에서 H_0를 기각.

단, 여기서 $\chi^2(p;\alpha)$는 자유도 p를 갖는 카이제곱분포의 상위 $100\alpha\%$ 백분위수이다.

3. 미니탭을 이용한 검정

이 절에서는 미니탭을 이용하여 두 다변량모집단의 평균벡터의 차를 검정하는 예를 제시한다. 본 논문에서는 Windows 95용 Release 12.1을 사용하여 프로그램을 작성한다. 미니탭과 그것의 사용법에 관한 내용은 MINITAB Inc.에서 발행한 사용안내서(1998)와 연세대학교 응용통계학과에서 저술한 사용해설서(1993)에 상세히 주어져 있다.

3.1 소표본의 경우

다음 예의 데이터는 강병서(1988, p.245)에서 인용한 데이터이다.

<예 3.1> 한국생화학연구소에서는 어떤 약이 뇌 속에서 세 가지 생화학성분의 수준에 변화를 일으킨다고 믿고 있다. 이것을 시험하기 위해 같은 종류의 쥐 24마리를 선택하여 두 그룹으로 나누었다. 각 그룹은 12마리씩이며 한 그룹은 통제그룹, 다른 그룹은 규칙적으로 약을 투여하는 약투여그룹으로 이름지었다. 통제그룹의 두 마리는 자연사했으나 두 그룹의 실험쥐들은 똑같은 먹이와 보살핌을 받았다. 얻어진 데이터는 <표 3.1>과 같다.

<표 3.1> <예 3.1>의 데이터

통 제			약 투 여		
A	B	C	A	B	C
1.21	0.61	0.70	1.40	0.50	0.71
0.92	0.43	0.71	1.17	0.39	0.69
0.80	0.35	0.71	1.23	0.44	0.70
0.85	0.48	0.68	1.19	0.37	0.72
0.98	0.42	0.71	1.38	0.42	0.71
1.15	0.52	0.72	1.17	0.45	0.70
1.10	0.50	0.75	1.31	0.41	0.70
1.02	0.53	0.70	1.30	0.47	0.67
1.18	0.45	0.70	1.22	0.29	0.68
1.09	0.40	0.69	1.00	0.30	0.70
			1.12	0.27	0.72
			1.09	0.35	0.73

두 모집단의 공분산행렬이 같은 정규분포의 가정 하에서 통제그룹과 약투여그룹의 두 평균벡터들이 같은지 여부를 유의수준 0.05에서 검정하고자 한다.

<그림 3.1>은 <표 3.1>의 데이터를 입력한 미니탭 워크시트(worksheet)이며, <그림 3.2>는 두 모집단의 평균벡터가 같은지 검정하기 위한 미니탭 프로그램과 출력결과이다. 즉 이 예에서는 2.1절의 δ_0를 영벡터로 놓아 귀무가설 $H_0 : \mu_1 = \mu_2$ 를 대립가설 $H_1 : \mu_1 \neq \mu_2$ 에 대비하여 검정하는 경우를 보인다.

<그림 3.1> <예 3.1>의 데이터를 입력한 미니탭 워크시트

Worksheet 1 ***

→	C1 통제A	C2 통제B	C3 통제C	C4 약투여A	C5 약투여B	C6 약투여C
1	1.21	0.61	0.70	1.40	0.50	0.71
2	0.92	0.43	0.71	1.17	0.39	0.69
3	0.80	0.35	0.71	1.23	0.44	0.70
4	0.85	0.48	0.68	1.19	0.37	0.72
5	0.98	0.42	0.71	1.38	0.42	0.71
6	1.15	0.52	0.72	1.17	0.45	0.70
7	1.10	0.50	0.75	1.31	0.41	0.70
8	1.02	0.53	0.70	1.30	0.47	0.67
9	1.18	0.45	0.70	1.22	0.29	0.68
10	1.09	0.40	0.69	1.00	0.30	0.70
11				1.12	0.27	0.72
12				1.09	0.35	0.73

<그림 3.2> <예 3.1>의 데이터에 대한 미니탭 프로그램과 출력결과

```
①  MTB > LET K21=10
②  MTB > LET K22=12
③  MTB > SET C20
   DATA> 0 0 0
   DATA> END
④  MTB > JOURNAL '3VAR-SMALL'
   Collecting keyboard input (commands and data) in file:3VAR-SMALL.MTJ
⑤  MTB > LET K1=MEAN(C1)
⑥  MTB > LET K2=MEAN(C2)
⑦  MTB > LET K3=MEAN(C3)
⑧  MTB > LET K4=MEAN(C4)
⑨  MTB > LET K5=MEAN(C5)
⑩  MTB > LET K6=MEAN(C6)
⑪  MTB > COPY K1-K3 C7
⑫  MTB > COPY K4-K6 C8
⑬  MTB > LET C9=C7-C8
⑭  MTB > NAME C7='XBAR1' C8='XBAR2' C9='XBAR1-XBAR2'
⑮  MTB > PRINT C7-C9

Data Display

 Row   XBAR1     XBAR2  XBAR1-XBAR2

  1    1.030   1.21500  -0.185000
```

```
      2   0.469   0.38833     0.080667
      3   0.707   0.70250     0.004500

⑯   MTB > COVARIANCE C1-C3 M1

⑰   MTB > COVARIANCE C4-C6 M2

⑱   MTB > NAME M1='S1' M2='S2'
⑲   MTB > PRINT M1 M2
```

Data Display

```
      Matrix S1

      0.0197556  0.0064667  0.0005556
      0.0064667  0.0056100  0.0001078
      0.0005556  0.0001078  0.0003567

      Matrix S2

      0.0139545  0.0061636 -0.0004409
      0.0061636  0.0054879 -0.0002682
     -0.0004409 -0.0002682  0.0002932

⑳   MTB > LET K31=K21-1
㉑   MTB > MULTIPLY M1 K31 M3
㉒   MTB > LET K32=K22-1
㉓   MTB > MULTIPLY M2 K32 M4
㉔   MTB > ADD M3 M4 M5
㉕   MTB > LET K7=1/(K21+K22-2)
㉖   MTB > MULTIPLY M5 K7 M6
㉗   MTB > NAME M6='S-POOLED'
㉘   MTB > PRINT M6
```

Data Display

```
      Matrix S-POOLED

      0.01656500  0.00630000  0.00000750
      0.00630000  0.00554283 -0.00009900
      0.00000750 -0.00009900  0.00032175

㉙   MTB > INVERT M6 M7
㉚   MTB > NAME M7='S-POOLED INVERSE'
㉛   MTB > PRINT M7
```

```
      Data Display

      Matrix S-POOLED INVERSE

        106.84   -122.16    -40.08
       -122.16    321.07    101.64
        -40.08    101.64   3140.21

㉜    MTB > LET C10=C9-C20
㉝    MTB > TRANSPOSE C10 M8
㉞    MTB > MULTIPLY M8 M7 M9
㉟    MTB > MULTIPLY M9 C10 K8
      Answer =          9.5961
㊱    MTB > LET K9=K21*K22/(K21+K22)*K8
㊲    MTB > NAME K9='T0-SQUARE'
㊳    MTB > LET K10=K21+K22-4
㊴    MTB > LET K11=K9*K10/((K21+K22-2)*3)
㊵    MTB > CDF K11 K12;
      SUBC> F 3, K10.
㊶    MTB > LET K13=1-K12
㊷    MTB > NAME K13='P-VALUE'
㊸    MTB > PRINT K9 K13

      Data Display

      T0-SQUARE      52.3421
      P-VALUE    0.000028813
㊹    MTB > NOJOURNAL
```

<그림 3.2>의 프로그램과 출력결과에 관한 설명은 다음과 같다. <그림 3.2>의 각 명령문에 붙여 놓은 번호를 따라 설명한다.

①~② 두 집단의 표본크기 n_1과 n_2를 지정한다.(K21, K22)

③ 귀무가설에서 주어진 벡터 $\underline{\delta}_0$를 지정한다.(C20)

④ 프로그램을 매크로(macro)로 만들어 저장하기 위한 명령문이다. ⑤~㊸까지의 명령문들이 '3VAR-SMALL.MTJ'라는 이름의 파일로 저장된다.

⑤~⑩ 각 그룹에 대하여 세 변수들의 표본평균을 구한다.(K1,K2,…,K6)

⑪ 통제그룹의 표본평균벡터 $\bar{X}_1$를 구한다.(C7)

⑫ 약투여그룹의 표본평균벡터 $\bar{X}_2$를 구한다.(C8)

⑬ $\bar{X}_1-\bar{X}_2$를 구한다.(C9)

⑭ C7, C8, C9에 이름을 붙인다.

⑮ $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_1-\bar{X}_2$를 출력한다.

⑯ 통제그룹의 표본공분산행렬 S_1을 구한다.(M1)

⑰ 약투여그룹의 표본공분산행렬 S_2를 구한다.(M2)

⑱ M1, M2에 이름을 붙인다.
⑲ S_1, S_2를 출력한다.
⑳~㉑ $(n_1-1)S_1$을 구한다.(M3)
㉒~㉓ $(n_2-1)S_2$를 구한다.(M4)
㉔ $(n_1-1)S_1+(n_2-1)S_2$를 구한다.(M5)
㉕~㉖ S_{pooled} 를 구한다.(M6)
㉗ M6에 이름을 붙인다.
㉘ S_{pooled} 를 출력한다.
㉙ S_{pooled} 의 역행렬 S_{pooled}^{-1} 를 구한다.(M7)
㉚ M7에 이름을 붙인다.
㉛ S_{pooled}^{-1} 를 출력한다.
㉜~㊱ 검정통계량 T_0^2의 값을 구한다.(K9)
㊲ K9에 이름을 붙인다.
㊳~㊶ 가설검정의 P값을 구한다.(K13)
㊷ K13에 이름을 붙인다.
㊸ T_0^2의 값과 P값을 출력한다.
㊹ 저장되는 프로그램의 끝을 나타내는 명령문이다.

위의 결과를 볼 때, 검정통계량 T_0^2 의 값은 52.3421 이고 P값은 0.000029이다. P값이 0.05보다 작으므로 유의수준 0.05에서 H_0를 기각한다. 즉 통제그룹과 약투여그룹의 모평균벡터가 같지 않다고 결론을 내릴 수 있다.

<예 3.2> <예 3.1>에서 매크로로 만들어진 프로그램은 표본크기와 귀무벡터를 바꿔가면서 또 다른 삼변량 문제에 얼마든지 이용될 수 있다. <예 3.1>의 실험에서 통제그룹에 6마리, 약투여그룹에 2마리의 쥐가 더 있다고 가정하고, 귀무가설

$$H_0 : \mu_1 - \mu_2 = \begin{pmatrix} -0.2 \\ 0.1 \\ 0.0 \end{pmatrix}$$

을 검정하고 싶다고 하자. <그림 3.3>은 데이터를 입력한 워크시트이며, <그림 3.4>는 위의 가설을 검정하기 위한 프로그램과 출력결과이다.

<그림 3.4>에 관한 설명은 다음과 같다.

①~② 두 집단의 표본크기 n_1과 n_2를 지정한다.(K21, K22)
③ 귀무가설에서 주어진 벡터 δ_0를 지정한다.(C20)
④ <예 3.1>에서 작성된 매크로 '3VAR-SMALL.MTJ'를 실행한다.

<그림 3.4>의 결과를 보면 검정통계량 T_0^2의 값은 2.35862 이고 P값은 0.543402이다. P값이 0.1보다 크므로, 유의수준 0.1에서 H_0를 기각할 수 없다.

<그림 3.3> <예 3.2>의 미니탭 워크시트

Worksheet 1 ***

→	C1 통제A	C2 통제B	C3 통제C	C4 약투여A	C5 약투여B	C6 약투여C
1	1.21	0.61	0.70	1.40	0.50	0.71
2	0.92	0.43	0.71	1.17	0.39	0.69
3	0.80	0.35	0.71	1.23	0.44	0.70
4	0.85	0.48	0.68	1.19	0.37	0.72
5	0.98	0.42	0.71	1 38	0.42	0.71
6	1.15	0.52	0.72	1.17	0.45	0.70
7	1.10	0.50	0.75	1.31	0.41	0.70
8	1.02	0.53	0 70	1.30	0.47	0.67
9	1.18	0.45	0.70	1.22	0.29	0.68
10	1.09	0.40	0.69	1.00	0.30	0.70
11	0.93	0.46	0.68	1.12	0.27	0.72
12	1.04	0.51	0.72	1.09	0.35	0.73
13	0.99	0.45	0.70	1.29	0.46	0.70
14	1.12	0.53	0.74	1.33	0.45	0.72
15	1.05	0 44	0.71			
16	0.89	0.41	0.69			

<그림 3.4> <예 3.2>의 데이터에 의한 미니탭 프로그램과 출력결과

```
①   MTB > LET K21=16
②   MTB > LET K22=14
③   MTB > SET C20
    DATA> -0.2 0.1 0
    DATA> END
④   MTB > EXECUTE '3VAR-SMALL.MTJ'
    Executing from file: 3VAR-SMALL.MTJ
```

Data Display

Row	XBAR1	XBAR2	XBAR1-XBAR2
1	1.02000	1.22857	-0.208571
2	0.46813	0.39786	0.070268
3	0.70688	0.70357	0.003304

Data Display

```
 Matrix S1

 0.0144000  0.0048867  0.0009000
 0.0048867  0.0040429  0.0003204
 0.0009000  0.0003204  0.0003696

 Matrix S2

 0.0130593  0.0060352 -0.0002484
 0.0060352  0.0052335 -0.0001687
-0.0002484 -0.0001687  0.0002709
```

Data Display

```
 Matrix S-POOLED

 0.01377755  0.00541990  0.00036684
 0.00541990  0.00459570  0.00009334
 0.00036684  0.00009334  0.00032376
```

Data Display

```
 Matrix S-POOLED INVERSE

   139.26   -161.98   -111.09
  -161.98    407.28     66.12
  -111.09     66.12   3195.56

Answer =         0.3159
```

Data Display

```
TO-SQUARE      2.35862
P-VALUE    0.543402
```

3.2 대표본의 경우

<예 3.3> Jolicoeur와 Mosimann(1960)은 거북의 크기와 모양의 관계를 연구하였다. 이 연구에서 암수 거북 각각 24마리를 뽑아 측정한 결과 <표 3.2>와 같은 데이터를 얻었다.

<표 3.2> <예 3.3>의 데이터

암 거 북			수 거 북		
길이	너비	높이	길이	너비	높이
98	81	38	93	74	37
103	84	38	94	78	35
103	86	42	96	80	35
105	86	42	101	84	39
109	88	44	102	85	38
123	92	50	103	81	37
123	95	46	104	83	39
133	99	51	106	83	39
133	102	51	107	82	38
133	102	51	112	89	40
134	100	48	113	88	40
136	102	49	114	86	40
138	98	51	116	90	43
138	99	51	117	90	41
141	105	53	117	91	41
147	108	57	119	93	41
149	107	55	120	89	40
153	107	56	120	93	44
155	115	63	121	95	42
155	117	60	125	93	45
158	115	62	127	96	45
159	118	63	128	95	45
162	124	61	131	95	46
177	132	67	135	106	47

암수 거북 모집단의 평균벡터가 같은지의 여부를 유의수준 0.05에서 검정하고자 한다. 두 모집단의 공분산행렬 Σ_1과 Σ_2는 같지 않다고 가정한다.

<표 3.2>의 데이터를 입력한 워크시트는 <그림 3.5>와 같으며, 이 문제를 해결하기 위한 미니탭 프로그램과 출력결과는 <그림 3.6>에 주어져 있다.

<그림 3.6>의 프로그램과 출력결과에 관한 설명은 다음과 같다.

①~② 두 집단의 표본크기 n_1과 n_2를 지정한다.(K21, K22)

③ 귀무가설에서 주어진 벡터 δ_0를 지정한다.(C20)

④ 프로그램을 매크로로 만들어 저장하기 위한 명령문이다. ⑤~㊱까지의 명령문들이 '3VAR-LARGE.MTJ'라는 이름의 파일로 저장된다.

⑤~⑩ 암수 거북 집단 각각에 대하여 세 변수들의 표본평균을 구한다. (K1,K2,…,K6)

⑪ 암거북 집단의 표본평균벡터 $\bar{X}_1$를 구한다.(C7)

⑫ 수거북 집단의 표본평균벡터 $\bar{X}_2$를 구한다.(C8)

<그림 3.5> <예3.3>의 데이터를 입력한 미니탭 워크시트

Worksheet 1 ***

↓	C1 암길이	C2 암너비	C3 암높이	C4 수길이	C5 수너비	C6 수높이
1	98	81	38	93	74	37
2	103	84	38	94	78	35
3	103	86	42	96	80	35
4	105	86	42	101	84	39
5	109	88	44	102	85	38
6	123	92	50	103	81	37
7	123	95	46	104	83	39
8	133	99	51	106	83	39
9	133	102	51	107	82	38
10	133	102	51	112	89	40
11	134	100	48	113	88	40
12	136	102	49	114	86	40
13	138	98	51	116	90	43
14	138	99	51	117	90	41
15	141	105	53	117	91	41
16	147	108	57	119	93	41
17	149	107	55	120	89	40
18	153	107	56	120	93	44
19	155	115	63	121	95	42
20	155	117	60	125	93	45
21	158	115	62	127	96	45
22	159	118	63	128	95	45
23	162	124	61	131	95	46
24	177	132	67	135	106	47

<그림 3.6> <예 3.2>의 데이터에 대한 미니탭 프로그램과 출력결과

```
①   MTB > LET K21=24
②   MTB > LET K22=24
③   MTB > SET C20
    DATA> 0 0 0
    DATA> END
④   MTB > JOURNAL '3VAR-LARGE'
    Collecting keyboard input (commands and data) in file:3VAR-LARGE.MTJ

⑤   MTB > LET K1=MEAN(C1)
⑥   MTB > LET K2=MEAN(C2)
⑦   MTB > LET K3=MEAN(C3)
⑧   MTB > LET K4=MEAN(C4)
```

```
⑨   MTB > LET K5=MEAN(C5)
⑩   MTB > LET K6=MEAN(C6)
⑪   MTB > COPY K1-K3 C7
⑫   MTB > COPY K4-K6 C8
⑬   MTB > LET C9=C7-C8
⑭   MTB > NAME C7='XBAR1' C8='XBAR2' C9='XBAR1-XBAR2'
⑮   MTB > PRINT C7-C9
```

Data Display

Row	XBAR1	XBAR2	XBAR1-XBAR2
1	136.042	113.375	22.6667
2	102.583	88.292	14.2917
3	52.042	40.708	11.3333

```
⑯   MTB > COVARIANCE C1-C3 M1

⑰   MTB > COVARIANCE C4-C6 M2

⑱   MTB > NAME M1='S1' M2='S2'
⑲   MTB > PRINT M1 M2
```

Data Display

Matrix S1

451.520	270.975	165.955
270.975	171.732	101.844
165.955	101.844	64.737

Matrix S2

138.766	79.147	37.375
79.147	50.042	21.654
37.375	21.654	11.259

```
⑳   MTB > LET K31=1/K21
㉑   MTB > MULTIPLY M1 K31 M3
㉒   MTB > LET K32=1/K22
㉓   MTB > MULTIPLY M2 K32 M4
㉔   MTB > ADD M3 M4 M5
```

```
㉕  MTB > INVERT M5 M6
㉖  MTB > NAME M6='INVERSE'
㉗  MTB > PRINT M6
```

Data Display

```
 Matrix INVERSE

  0.86126  -0.80477  -0.99652
 -0.80477   1.89039  -0.91879
 -0.99652  -0.91879   4.47508
```

```
㉘  MTB > LET C10=C9-C20
㉙  MTB > TRANSPOSE C10 M7
㉚  MTB > MULTIPLY M7 M6 M8
㉛  MTB > MULTIPLY M8 C10 K8
    Answer =        72.3816
㉜  MTB > NAME K8='TO-SQUARE'
㉝  MTB > CDF K8 K9;
    SUBC> CHISQUARE 3.
㉞  MTB > LET K10=1-K9
㉟  MTB > NAME K10='P-VALUE'
㊱  MTB > PRINT K8 K10
```

Data Display

```
    TO-SQUARE     72.3816
    P-VALUE    0.000000000
㊲  MTB > NOJOURNAL
```

⑬ $\bar{X}_1 - \bar{X}_2$를 구한다.(C9)

⑭ C7, C8, C9에 이름을 붙인다.

⑮ $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_1 - \bar{X}_2$를 출력한다.

⑯ 암거북 집단의 표본공분산행렬 S_1을 구한다.(M1)

⑰ 수거북 집단의 표본공분산행렬 S_2를 구한다.(M2)

⑱ M1, M2에 이름을 붙인다.

⑲ S_1, S_2를 출력한다.

⑳~㉑ $\frac{1}{n_1} S_1$을 구한다.(M3)

㉒~㉓ $\frac{1}{n_2} S_2$를 구한다.(M4)

㉔ $\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2$를 구한다.(M5)

㉕ $\left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2\right)^{-1}$를 구한다.(M6)

㉖ M6에 이름을 붙인다.

㉗ $\left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2\right)^{-1}$를 출력한다.

㉘~㉛ 검정통계량 T_0^2의 값을 구한다.(K8)

㉜ K8에 이름을 붙인다.

㉝~㉞ 가설검정의 P값을 구한다.(K10)

㉟ K10에 이름을 붙인다.

㊱ T_0^2의 값과 P값을 출력한다.

㊲ 저장되는 프로그램의 끝을 나타내는 명령문이다.

위의 결과를 볼 때, T_0^2 의 값은 72.3816이고 P값은 0.000000이다. P값이 0.05보다 작으므로 유의수준 0.05에서 H_0를 기각한다. 즉 암거북 집단과 수거북 집단의 모평균벡터가 다르다고 결론을 내릴 수 있다.

<예 3.4> <예 3.3>에서 작성된 프로그램도 표본크기와 귀무벡터를 바꿔가면서 대표본의 또 다른 삼변량 문제에 간편히 이용될 수 있다. <예 3.3>에서 암거북 1마리, 수거북 2마리가 더 있다고 가정하고, 귀무가설

$$H_0 : \mu_1 - \mu_2 = \begin{pmatrix} 22 \\ 15 \\ 12 \end{pmatrix}$$

을 검정하고 싶다고 하자. <그림 3.7>은 데이터를 입력한 워크시트이며, <그림 3.8>은 위의 가설을 검정하기 위한 프로그램과 출력결과이다.

<그림 3.8>에 관한 설명은 다음과 같다.

①~② 두 집단의 표본크기 n_1과 n_2를 지정한다.(K21, K22)

③ 귀무가설에서 주어진 벡터 δ_0를 지정한다.(C20)

④ <예 3.3>에서 작성된 매크로 '3VAR-LARGE.MTJ'를 실행한다.

<그림 3.8>의 결과를 보면 검정통계량 T_0^2의 값은 3.16149이고 P값은 0.367391이다. P값이 0.1보다 크므로, 유의수준 0.1에서 H_0를 기각할 수 없다.

<그림 3.7> <예 3.4>의 미니탭 워크시트

Worksheet 1 ***

	C1	C2	C3	C4	C5	C6
↓	앞:길이	앞:너비	앞:높이	수:길이	수:너비	수:높이
1	98	81	36	93	74	37
2	103	84	38	94	78	35
3	103	86	42	96	80	35
4	105	86	42	101	84	39
5	109	88	44	102	85	38
6	123	92	50	103	81	37
7	123	95	46	104	83	39
8	133	99	51	106	83	39
9	133	102	51	107	82	38
10	133	102	51	112	99	40
11	134	100	48	113	98	40
12	136	102	49	114	86	40
13	138	98	51	116	90	43
14	138	99	51	117	90	41
15	141	105	53	117	91	41
16	147	108	57	119	93	41
17	149	107	55	120	89	40
18	153	107	56	120	93	44
19	155	115	63	121	95	42
20	155	117	60	125	93	45
21	158	115	62	127	96	45
22	159	118	63	128	95	45
23	162	124	61	131	95	46
24	177	132	67	135	106	47
25	179	133	67	138	104	48
26				140	107	46

<그림 3.8> <예 3.4>의 데이터에 대한 미니탭 프로그램과 출력결과

```
①  MTB > LET K21=25
②  MTB > LET K22=26
③  MTB > SET C20
   DATA> 22 15 12
   DATA> END
④  MTB > EXECUTE '3VAR-LARGE.MTJ'
   Executing from file: 3VAR-LARGE.MTJ

   Data Display

    Row     XBAR1     XBAR2  XBAR1-XBAR2

      1    137.76   115.346      22.4138
      2    103.80    89.615      14.1846
      3     52.64    41.192      11.4477
```

```
Data Display

 Matrix S1

  506.523   311.950   184.743
  311.950   201.583   115.800
  184.743   115.800    70.990

 Matrix S2

  176.235   105.498    46.211
  105.498    68.086    27.797
   46.211    27.797    13.362

Data Display

 Matrix INVERSE

   0.93783  -0.90118  -1.03159
  -0.90118   1.87618  -0.72615
  -1.03159  -0.72615   4.35262

Answer =          3.1615

Data Display

T0-SQUARE        3.16149
P-VALUE      0.367391
```

4. 결 론

본 논문에서는 두 다변량모집단의 평균벡터의 차에 관한 가설을 검정하는 문제를 소표본의 경우와 대표본의 경우로 나누어 고려하였다. 제3절에서 볼 수 있듯이, 통계패키지 미니탭을 이용하면 복잡한 계산을 컴퓨터로 처리하여 원하는 결과를 간편히 얻을 수 있다. 다변량 문제의 경우는 벡터와 행렬을 많이 다루게 되므로 미니탭의 행렬 연산 기능이 특히 중요한 역할을 한다.

본 논문에서는 명령문 방식으로 작업을 하는 과정을 보였다. 메뉴 방식을 이용해서도 동일한 작업을 손쉽게 처리할 수 있으나, 그 과정을 보이는 것은 지면을 너무 많이 차지하므로 생략하였다. 또한 본 논문에서는 삼변량의 경우를 예로 들었지만, 제시된 프로그램 기법은 임의의 p변량모집단의 경우로 그대로 확장될 수 있다는 것이 명백하다. <예 3.1>부터 <예 3.4>까지에 걸쳐서 보였듯이 작성된 프로그램은 매크로로 만들어져 저장되어 필요할 때마다 불리어서 사용될 수 있다.

참 고 문 헌

[1] 강병서(1988), "다변량통계학", 법문사.

[2] 연세대학교 응용통계학과(1993), "미니탭"(개정증보판), 자유아카데미.

[3] Anderson, T. W.(1984), *An Introduction to Multivariate Statistical Analysis*(2nd ed.), Wiley.

[4] Johnson, R. A. and Wichern, D. W.(1982), *Applied Mulivariate Statistical Analysis*, Prentice-Hall.

[5] Jolicoeur, P. and Mosimann, J. E.(1960), "Size and Shape Variation in the Painted Turtle : A Principal Component Analysis", *Growth*, 24, 339-354.

[6] MINITAB Inc.(1998), *MINITAB User's Guide 1&2*.