

다변량분석 중간고사(2012) 학번:\_\_\_\_\_ 이름:\_\_\_\_\_

[1] 다음은 여러 종류의 Cereal 30g 에 포함된 영양소 함량(mg)을 측정한 자료와 이와 관련된 SAS output 이다.

ID	Brand	sugar (y1)	fat (y2)	sodium (y3)
1	0	10	0	170
2	0	2	0	270
3	0	2	0	300
4	0	2	2	280
5	1	13	1	210
6	1	11	0	180
7	1	12	1.5	200
8	1	9	2.5	200
9	1	11	0.5	220
10	1	2	0	330

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
sugar	10	7.40000	4.76562	74.00000	2.00000	13.00000
fat	10	0.75000	0.95015	7.50000	0	2.50000
sodium	10	236.00000	54.81281	2360	170.00000	330.00000

1.1 위 자료에서 공분산  $S_{23} = \text{Cov}(y_2, y_3)$  을 공식을 이용하여 계산하라(과정포함) (4점)

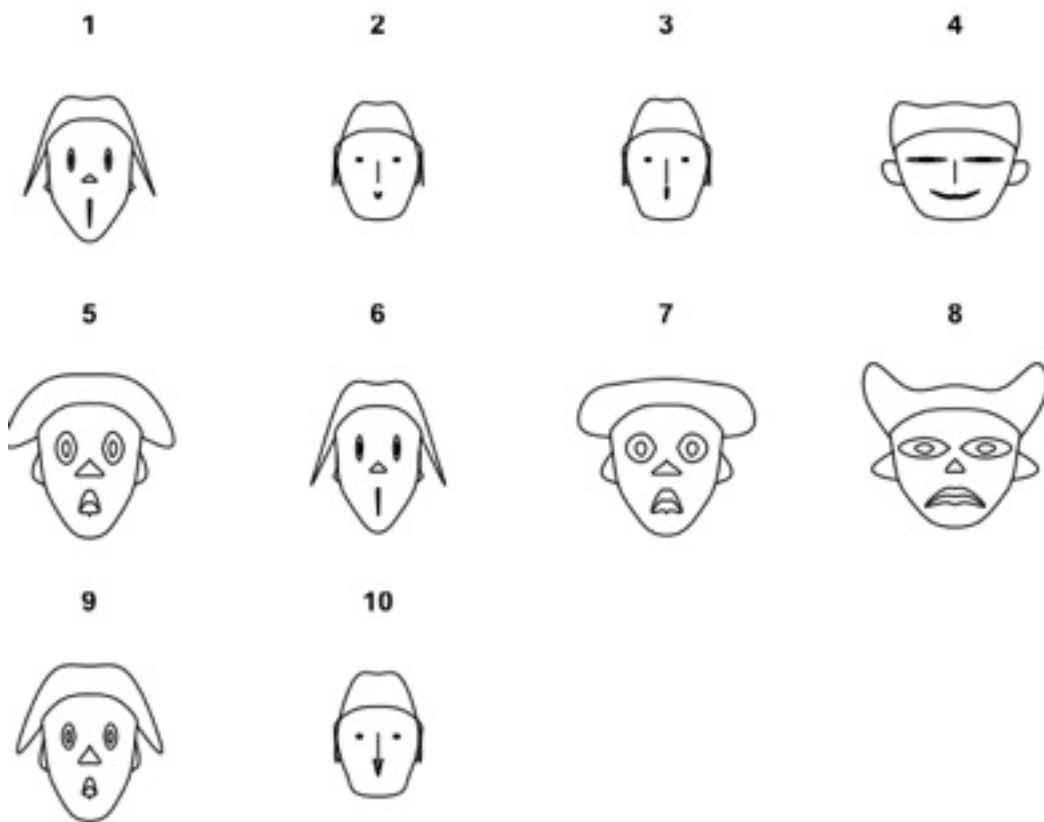
1.2 확률벡터  $y$ 를 아래와 같이 정의할 때,  $y$ 가 다변량 정규분포  $N_3(\mu, \Sigma)$  를 따른다고 가정하자. 이 때, 다변량 정규분포의 확률밀도함수  $f(y; \mu, \Sigma)$  를 완성하기 위한 A,B,C, 값을 적어보아라. (각 2점)

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad f(y; \mu, \Sigma) = (2\pi)^A |\Sigma|^B \exp(C)$$

1.3 확률벡터  $y \sim N_3(\mu, \Sigma)$  일 때, 상수행렬  $A$ 에 대하여  $Ay$ 의 분포는  $N_3(M, V)$ 를 따른다. 이 때  $M, V$ 를 각각 써 보아라. (6점)

1.4 확률벡터  $y \sim N_3(\mu, \Sigma)$  일 때,  $(y - \mu)^T \Sigma^{-1} (y - \mu)$ 는 어떤 분포를 가지는지 적어 보아라 (4점)

1.5 아래는 위 자료를 Chernoff의 얼굴그림 faces(.)함수로 그린 것이다.



크게 두 그룹으로 나눌 때 각 그룹에 속한 ID를 적어보고 각 그룹의 특징을 간략하게 적어 보아라. (10점)

1.6 위 자료는 두 개의 회사(Brand=1, 0)에 대한 측정 결과이다. 두 회사의 제품이 서로 함량 차이가 존재하는지 알아보기 위한 SAS 코드를 적어보아라.

(10점)

Proc GLM;

\_\_\_\_\_ ;

Model \_\_\_\_\_ ;

\_\_\_\_\_;

Run;

1.7 Brand의 변화(B= 0에서 1로)에 의한 sugar 함량의 변화는 회귀분석 모형식으로

$$y_{1i} = \beta_{01} + \beta_{11}B_i + \varepsilon_{1i}$$

라고 표현할 수 있다. 같은 방법으로 fat 과 sodium 도 아래와 같은 회귀모형식이 가능하다.

$$y_{2i} = \beta_{02} + \beta_{12}B_i + \varepsilon_{2i}$$

$$y_{3i} = \beta_{03} + \beta_{13}B_i + \varepsilon_{3i}$$

이와 같이, 3개의 반응변수에 대한 회귀식을 행렬로 표현하려고 한다. 아래를 완성하여라.

(5점)

$$\begin{pmatrix} y_{11} & y_{21} & y_{31} \\ y_{12} & y_{22} & y_{32} \\ \vdots & \vdots & \vdots \\ y_{110} & y_{210} & y_{310} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{pmatrix} + E$$

1.8 위 회귀모형식을 분석하기 위해 아래와 같은 SAS 코드를 작성하였다. Brand 별로 sugar 와 fat 의 차이가 동일한지를 분석하기 위한 코드를 완성하여라.

(5점)

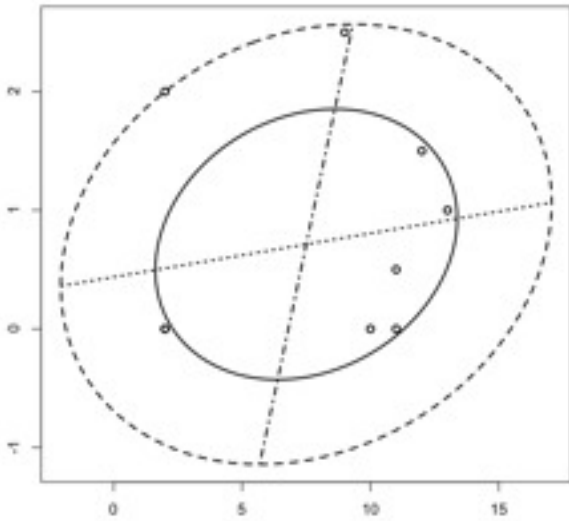
Proc Reg;

Model y1 y2 y3 = Brand ;

\_\_\_\_\_ ;

Run;

1.9 위 자료에서  $y_1=\text{sugar}$  와  $y_2=\text{fat}$  에 대한 `bvbox( . )` 함수를 실행한 결과 아래와 같은 그림이 나왔다. 이 그림을 보고 `sugar` 와 `fat`의 분포에 대해 아는대로 적어 보시오.  
(10점)



1.10 위 자료에서  $y_1$  (`sugar`),  $y_2$  (`fat`),  $y_3$  (`sodium`)의 부분상관계수와 상관계수를 구해 보았더니 아래와 같았다.

$$\phi_{12.3} = 0.0317 \quad r_{12} = 0.184$$

이를 이용하여 `sugar`와 `fat` 간의 상관관계에 대해 논하라.  
(10점)

[2] 확률벡터  $y_i \sim N_3(\mu, \Sigma)$ ,  $i = 1, 2, \dots, n$  일 때,  $T_0^2 = (\bar{y} - \mu)^T S^{-1}(\bar{y} - \mu)$  에 대하여

$$C \cdot T_0^2 \sim F_{v1, v2}$$

의 분포를 따른다고 할 때,

2.1  $C$  의 값을 적어라  
(4점)

2.2  $F$ 분포의  $v1, v2$  값을 각각 적어라.  
(6점)

2.3 이 통계량을 무엇이라고 하는가?  
(5점)