STA 135
Winter 2020
Sample Final

Name: ______________________

Time Limit: 120 Minutes

This exam contains 6 pages (including this cover page) and 5 questions. Total of points is 100.

Answers without supporting work will not be given credit, unless otherwise stated. All answers should be completely simplified. You can take four pieces (eight sides) of A4 paper as cheatsheets. Only plain calculators are allowed. No graphing calculators or other electronic devices may be used on exams.

Grade Table (for teacher use only)

Question	Points	Score
1	20	
2	20	
3	20	
4	20	
5	20	
Total:	100	

---

1. (25 points) Consider a sample of size $n_1 = 18$ from $\mathcal{N}_2(\vec{\mu}_1, \mathbf{\Sigma}_1)$ and a sample of size $n_2 = 18$ from $\mathcal{N}_2(\vec{\mu}_2, \mathbf{\Sigma}_2)$. Assume $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. The summary statistics for these two samples are

$$\overline{\vec{x}}_1 = \begin{bmatrix} 80 \\ 83 \end{bmatrix}, \quad \boldsymbol{S}_1 = \begin{bmatrix} 16 & 8 \\ 8 & 16 \end{bmatrix}, \quad \overline{\vec{x}}_2 = \begin{bmatrix} 83 \\ 80 \end{bmatrix}, \quad \boldsymbol{S}_2 = \begin{bmatrix} 16 & 8 \\ 8 & 16 \end{bmatrix}.$$

   1. Test $H_0 : \vec{\mu}_1 = \vec{\mu}_2$ at the level of $\alpha = .05$ with Hotelling's T^2;
   2. Plot 95% confidence region for $\vec{\mu}_1 - \vec{\mu}_2$;
   3. Find $\geq$ 95% simultaneous confidence intervals for $\mu_{1j} - \mu_{2j}$, $j = 1, 2$ with Bonferroni correction.

2. (25 points) Suppose we are given the data

z_1 :	2	2	2	0	-1	-2	-3
z_2 :	-1	-2	-3	0	2	2	2
y :	1	0	1	0	-1	0	-1

We aim at fitting the linear model $Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \epsilon_i, \quad i = 1, 2, \ldots, 7$.
(1) Find the least square estimate $\hat{\vec{\beta}}$, the R^2 statistic, $\hat{\sigma}^2$ and $\widehat{\text{Cov}}(\vec{\beta})$;
(2) Find a 95% confidence interval for β_1;
(3) Test $H_0 : \beta_1 = \beta_2 = 0$ at the level of $\alpha = 0.05$;
(4) Find a 95% prediction interval for a new response Y_0 corresponding to $(\bar{z}_1, \bar{z}_2)$, where $\bar{z}_j$ is the sample mean of z_j, $j = 1, 2$.

3. (25 points) You are given a sample:

$$\begin{array}{llllllll} X_1: & 2 & 2 & 2 & 0 & -1 & -2 & -3 \\ X_2: & -1 & -2 & -3 & 0 & 2 & 2 & 2. \end{array}$$

(a) Give the formula for the first and second sample principal components $\widehat{Y}_1$ and $\widehat{Y}_2$;
(b) Determine the proportion of total sample variance due to the first sample principal component;
(c) Compare the contributions of the two variates to the determination of the first sample principal component based on loadings;
(d) Compare the contributions of the two variates to the determination of the first sample principal component based on sample correlations.

4. (15 points) Consider three independent samples from three classes:
Class 1: Sample size n_1, sample mean $\bar{\vec{x}}_1$, sample covariance $\boldsymbol{S}_1$;
Class 2: Sample size n_2, sample mean $\bar{\vec{x}}_2$, sample covariance $\boldsymbol{S}_2$;
Class 3: Sample size n_3, sample mean $\bar{\vec{x}}_3$, sample covariance $\boldsymbol{S}_3$.

We use the pooled sample covariance matrix

$$\boldsymbol{S}_{pooled} = \frac{1}{n_1 + n_2 + n_3 - 3} \left((n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2 + (n_3 - 1)\boldsymbol{S}_3\right)$$

to classify the three classes. If the classification boundary between Class 1 and Class 2 is denoted as L_{12}, the boundary between Class 2 and Class 3 is L_{23}, and the boundary between Class 3 and Class 1 is L_{31}. Show that if L_{12} and L_{23} intersects at the point P, then L_{31} also goes through P.

5. (10 points) Suppose we are given the data

$$\begin{array}{r|cccc} X_1: & x_{11} & x_{21} & \dots & x_{n1} \\ X_2: & x_{12} & x_{22} & \dots & x_{n2} \\ \hline Y: & y_1 & y_2 & \dots & y_n \end{array}$$

We assume the data has already been centered, i.e., $\bar{y} = \bar{x}_1 = \bar{x}_2 = 0$. Denote the first and second sample principal components of X_1 and X_2 as $\widehat{Z}_1$ and $\widehat{Z}_2$ with the observation

$$\begin{array}{rcccc} \widehat{Z}_1: & \hat{z}_{11} & \hat{z}_{21} & \dots & \hat{z}_{n1} \\ \widehat{Z}_2: & \hat{z}_{12} & \hat{z}_{22} & \dots & \hat{z}_{n2}. \end{array}$$

Consider the R^2 statistics of the following linear regression (We don't need to include the slope since all data are centered):

- $y_i = \beta_1 \hat{z}_{i1} + \epsilon_i$, R^2 statistic is denoted as R_1^2;
- $y_i = \beta_2 \hat{z}_{i2} + \epsilon_i$, R^2 statistic is denoted as R_2^2;
- $y_i = \beta_1 \hat{z}_{i1} + \beta_2 \hat{z}_{i2} + \epsilon_i$, R^2 statistic is denoted as R_{full}^2.

Show that $R_{full}^2 = R_1^2 + R_2^2$.
(Hint: First show that in each regression, the explained sum of squares can be written as $\vec{y}^\top \boldsymbol{H} \vec{y}$, where $\boldsymbol{H}$ is the hat matrix corresponding to that particular regression.)