JARROD
CLEMONS

STAT 520 FINAL EXAM, SPRING 2021

1. CLUSTERING IS A METHOD OF ANALYSIS IN WHICH WE GROUP VARIABLES TOGETHER ACCORDING TO SIMILARITY TO ONE ANOTHER. THIS PROCESS RESULTS IN SEVERAL GROUPS CALLED "CLUSTERS".

2. THE DIFFERENCE BETWEEN REGRESSION AND DISCRIMINANT ANALYSIS IS THAT REGRESSION AIMS TO FIND THE CORRELATION BETWEEN TWO VARIABLES, A RESPONSE AND A PREDICTOR, WHEREAS DISCRIMINANT ANALYSIS USES A CATEGORICAL VARIABLE AS THE RESPONSE, USING A SET OF PREDICTORS TO SEPERATE OUTCOME GROUPS.

3a MULTIVARIATE REGRESSION

b NO, THE CHI-SQUARE TEST FOR HOMOGENEITY HAS A P-VALUE $< .0001$, REJECTING THE NOTION OF HOMOGENEITY.

c ASSUMPTIONS ARE HOMOGENEITY OF VAR-COV MATRIX TO DETERMINE POOL STATUS.

4a MULTIPLE REGRESSION

b $\underset{(110,3)}{Y} = \underset{(110,4)}{X}\,\underset{(4,3)}{\beta} + \underset{(110,3)}{\varepsilon}$; $n = 110 \quad p = 3 \quad k = 3$

ASSUMPTIONS INCLUDE: MULTIVARIATE NORMAL DISTRIBUTION
HOMOGENEITY OF VAR-COV MATRIX

c $H_0: \underline{\beta}_1 = \underline{\beta}_2 = \underline{\beta}_3 = 0$

$H_1: \underline{\beta}_1 \neq \underline{\beta}_2 \neq \underline{\beta}_3 \neq 0$

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \beta_{03} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \quad m = I_4$$

d $H_0: (\beta_{31} - \beta_{32}) - (\beta_{31} - \beta_{33}) = 0$, $H_1: \beta_{31} \neq \beta_{32} \neq \beta_{33}$

$$L = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \quad m = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}$$

e $H_0: (\beta_{11} - \beta_{21}) - (\beta_{12} - \beta_{22}) = 0$
$(\beta_{11} - \beta_{21}) - (\beta_{13} - \beta_{23}) = 0$

$H_1: (\beta_{11} - \beta_{21}) \neq (\beta_{12} - \beta_{22}) \neq (\beta_{13} - \beta_{23})$

5 a MANOVA

b. $\underline{Y}_{ij} = \underline{\mu} + \underline{\tau}_i + \underline{\varepsilon}_{ij}$; $i = 1-t$, $j = 1-n_i$

c Assumptions include: multivariate normality and homogeneity of var-cov matrix.

d $H_0: \underline{\tau}_1 = \underline{\tau}_2 = \underline{\tau}_3 = 0$ or $H_0: \underline{\tau}_1 - \underline{\tau}_2 = 0, \underline{\tau}_1 - \underline{\tau}_3 = 0$

$H_1: \underline{\tau}_1 \neq \underline{\tau}_2 \neq \underline{\tau}_3$

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \\ \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{bmatrix} \quad M = I_4$$

e. $H_0: \mu_1 = \mu_2 = \mu_3 = 0$ or $H_0: \mu_1 - \mu_2 = 0, \mu_1 - \mu_3 = 0$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \quad M = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \quad H_1: \mu_1 \neq \mu_2 \neq \mu_3$$

f.
```
PROC GLM data = a;
CLASS SPECIES;
MODEL LENGTH1 LENGTH2 LENGTH3 = SPECIES/NOUNI;
MANOVA L = INTERCEPT M = LENGTH1 - LENGTH2, LENGTH1 - LENGTH3;
QUIT;
```