一．简答 （每题5分）

1. 定义欧氏距离和曼哈顿距离；

2. 解释碎石图（scree plot）的含义；

3. 解释层次聚类和逐步聚类的优缺点；

4. 解释主成分分析和因子分析的原理和分别。

二．（15分）令**X**1 *j* 和**X** 2 *j* 为成对对比实验中的治疗方法1和方法2的响应变量*j*1,2,…,*n.* 假设**D** *j*  **X**1 *j*  **X**2 *j* 服从多元正态分布，计算30个样本的均值和方差分别为已知

检验在5%显著水平下是否存在治疗效果差异，并计算两种方法均值差异的95%联合置信域（写出不等式）。qf(c(0.05,0.95),3,27) = 0.1159239, 2.9603513

三． （15分）假设200个随机变量*X* 1 , *X* 2 , *X* 3 观测值的样本协方差阵为，求样本主成分*Y*1  , *Y*2, *Y*3和它们解释原变量方差的比例。

四． 令*Z*1, *Z*2, *Z*3 为三维标准化随机变量，其协方差阵为，  为单因子模型，且有  **,**

1. 填空 （9分）

2. 求变量的共同度，写出并验证因子模型 （6分）

3. 求因子F1解释的原变量方差的比例（5分）

五． （15分）令8个数据点的距离矩阵如下，试用最短距离法（single linkage)做系统聚类，并画出谱系图（dendogram）.



六．下列是一判别分析的输出结果，解释变量为年龄和收入，响应变量为是否有忧郁倾向。

llda<-lda(CASES~AGE+INCOME,lad)

llda

Call:

lda(CASES ~ AGE + INCOME, data = lad)

Prior probabilities of groups:

0 1

0.5 0.5

Group means:

AGE INCOME

0 55.42857 19.07143

1 36.07143 14.14286

Coefficients of linear discriminants:

LD1

AGE -0.04948687

INCOME -0.03859237

1. 写出判别方程（10分）
2. 现有一个新数据AGE＝30，INCOME=8, 根据判别方程判断他属于哪组。（5分）