

姓名：
李金哲

学号：
201707010119

班级：
信 171

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。

北京建筑大学 2019/ 2020 学年 第 2 学期 考试

数据分析 课程期末 答题纸 B

一、
1)

表 1-1-1 均值

| 报告 | | | |
|-----|---------|--------|---------|
| | HbA1C | HDLC | LDLC |
| 平均值 | 6.6027 | 1.1747 | 4.3618 |
| 个案数 | 110 | 110 | 110 |
| 标准差 | 1.21968 | .26734 | 1.38226 |

由表 1-1-1，可以得到 HbA1C 的平均值为 6.6027， HDLC 的平均值为 1.1747， LDLC 的平均值为 4.3618

2)

表 1-2-1 方差、标准差、偏度、峰度

| 描述统计 | | | | | | | | |
|---------------|-----|--------|---------|-------|-------|------|-------|------|
| | 个案数 | 平均值 | 标准差 | 方差 | 偏度 | | 峰度 | |
| | | | | | | 标准 | | 标准 |
| | 统计 | 统计 | 统计 | 统计 | 统计 | 误差 | 统计 | 误差 |
| HbA1C | 110 | 6.6027 | 1.21968 | 1.488 | 1.031 | .230 | .395 | .457 |
| HDLC | 110 | 1.1747 | .26734 | .071 | -.004 | .230 | -.478 | .457 |
| LDLC | 110 | 4.3618 | 1.38226 | 1.911 | .033 | .230 | -.648 | .457 |
| 有效个案数 (成列) | 110 | | | | | | | |

由表1-2-1，HbA1C 的方差为 1.488 、标准差为 1.21968 、偏度为 1.031、峰度为 0.395；
HDLC的方差为 0.071 、标准差为 0.26734 、偏度为 -0.004、峰度为 -0.478；
LDLC的方差为 1.911 、标准差为 1.38226 、偏度为 0.033、峰度为 -0.648

3)

表 1-3-1 中位数、上下四分位数

| 统计 | | |
|-------|----|--------|
| HbA1C | | |
| 个案数 | 有效 | 110 |
| | 缺失 | 0 |
| 中位数 | | 6.3000 |
| 百分位数 | 25 | 5.7000 |
| | 50 | 6.3000 |
| | 75 | 7.6000 |

由表 1-3-1，可以得到 HbA1C 的中位数为 6.3000、上四分位数为 5.7000，下四分位数为 7.6000.

4)

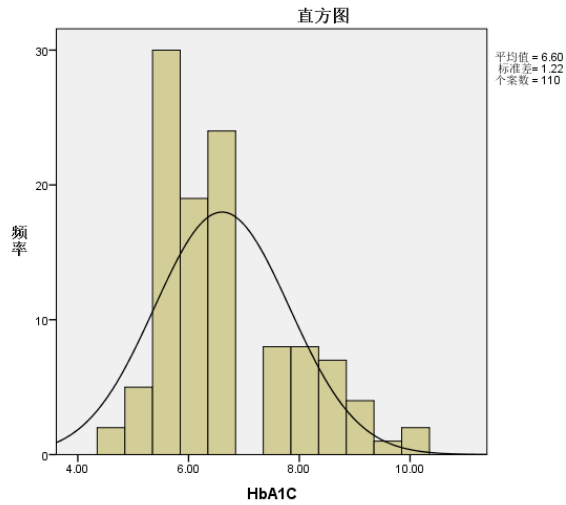


图1-4-1 直方图

5)

表1-5-1 相关性矩阵

| | | 相关性 | | |
|-------|---------|-------|-------|-------|
| | | HbA1C | HDLC | LDLC |
| HbA1C | 皮尔逊相关性 | 1 | -.179 | -.067 |
| | 显著性（双尾） | | .061 | .485 |
| | 个案数 | 110 | 110 | 110 |
| HDLC | 皮尔逊相关性 | -.179 | 1 | -.132 |
| | 显著性（双尾） | .061 | | .170 |
| | 个案数 | 110 | 110 | 110 |
| LDLC | 皮尔逊相关性 | -.067 | -.132 | 1 |
| | 显著性（双尾） | .485 | .170 | |
| | 个案数 | 110 | 110 | 110 |

由表1-5-1，可以得出假设概率p值为0.05，则 HbA1C 、HDLC、LDLC 三者的相关显著性均大于0.05，则说明三者没有线性关系。

北京建筑大学 2019/ 2020 学年 第 2 学期 考试

数据分析 课程期末 答题纸

姓名：
李金哲

学号：
201707010119

班级：
信 171

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。

一、
1)

表 2-1-1 主体间效应检验

| 主体间效应检验 | | | | | |
|-----------------|-----------------------|-----|-----------|------------|-------|
| 因变量： time | | | | | |
| 源 | III 类平方和 | 自由度 | 均方 | F | 显著性 |
| 修正模型 | 373. 105 ^a | 8 | 46. 638 | 774. 910 | . 000 |
| 截距 | 1857. 610 | 1 | 1857. 610 | 30864. 905 | . 000 |
| repairer | 220. 020 | 2 | 110. 010 | 1827. 858 | . 000 |
| type | 123. 660 | 2 | 61. 830 | 1027. 329 | . 000 |
| repairer * type | 29. 425 | 4 | 7. 356 | 122. 227 | . 000 |
| 误差 | 1. 625 | 27 | . 060 | | |
| 总计 | 2232. 340 | 36 | | | |
| 修正后总计 | 374. 730 | 35 | | | |

a. R 方 = .996 (调整后 R 方 = .994)

由于 repairer*type 的 F 统计量对饮的概率 P-值近似为 0，小于显著性水平 0. 05，因此不应拒绝原假设，认为不同剂量混合对病情缓解时间产生了显著的交互作用。

2)

表 2-2-1 repairer 表

1. repairer

| 因变量： time | | | | |
|-----------|--------|-------|----------|---------|
| repairer | 平均值 | 标准误差 | 95% 置信区间 | |
| | | | 下限 | 上限 |
| 1 | 3. 883 | . 289 | 3. 294 | 4. 473 |
| 2 | 7. 833 | . 289 | 7. 244 | 8. 423 |
| 3 | 9. 833 | . 289 | 9. 244 | 10. 423 |

表2-2-2 type表

2. type

| 因变量： time | | | | |
|-----------|--------|-------|----------|--------|
| type | 平均值 | 标准误差 | 95% 置信区间 | |
| | | | 下限 | 上限 |
| 1 | 4. 633 | . 289 | 4. 044 | 5. 223 |
| 2 | 7. 933 | . 289 | 7. 344 | 8. 523 |
| 3 | 8. 983 | . 289 | 8. 394 | 9. 573 |

根据表 2-2-1 和表 2-2-2 可知 repairer 类别 1 的置信区间 (3.294102, 4.472565)，类别 2 的置信区间 (7.244102, 8.422565)，类别 3 的置信区间 (9.244102, 10.422565)；

type 类别 1 的置信区间(4.044102, 5.222565), 类别 2 的置信区间(7.344102, 8.522565), 类别 3 的置信区间(8.394102, 9.572565)

姓名：
李金哲

学号：
201707010119

班级：
信 171

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。

北京建筑大学 2019/ 2020 学年 第 2 学期 考试

数据分析 课程期末 答题纸

三、

1)

由表3-1-1可知，回归方程显著性检验的 F 统计量的观测值为 87.134，其对应的概率 P值近似为 0。若显著性水平 a 为 0.05，因概率 P 值小于 a，拒绝回归方程显著性检验的原假设，即回归系数不同时为 0，解释变量与被解释变量间存在显著的线性关系，选择线性模型具有合理性。

根据表3-1-2，可以得到回归方程为：

$$y = - 18.895 * 0.158 * X1 + 0.011 * X2$$

但是若显著性水平a为0.05，常数的概率p值0.006、X1的概率p值0.001、X2的概率p值0.004，均小于a。故可以认为常数、X1、X2接受原假设。能够判断该回归方程是最优解

表3-1-1 检验结果

| ANOVA ^a | | | | | | |
|--------------------|----|-----------|-----|----------|--------|-------------------|
| 模型 | | 平方和 | 自由度 | 均方 | F | 显著性 |
| 1 | 回归 | 12204.035 | 2 | 6102.017 | 87.134 | .000 ^b |
| | 残差 | 1190.515 | 17 | 70.030 | | |
| | 总计 | 13394.550 | 19 | | | |

a. 因变量: Y

b. 预测变量: (常量), X2, X1

表3-1-2 系数表

| 模型 | | 未标准化系数 | | 标准化系数 | B 的 95.0% 置信区间 | | | |
|----|------|---------|-------|-------|----------------|------|---------|--------|
| | | B | 标准误差 | Beta | t | 显著性 | 下限 | 上限 |
| 1 | (常量) | -18.895 | 5.973 | | -3.163 | .006 | -31.496 | -6.293 |
| | X1 | .158 | .042 | .530 | 3.796 | .001 | .070 | .246 |
| | X2 | .011 | .003 | .460 | 3.294 | .004 | .004 | .019 |

a. 因变量: Y

2)

根据表3-1-2，可以得到线性方程系数的置信度为95%的置信区间，常量的置信区间为（ -31.496 - -6.293），X1的置信区间为（0.070 - 0.246），X2的置信区间为（0.004 - 0.019）

3)

| | | | | | | |
|-----------|----|-----|------|----------|----------|----------|
| 24: PRE_1 | | | | | | |
| | Y | X1 | X2 | PRE_1 | LMCI_1 | UMCI_1 |
| 19 | 70 | 268 | 3590 | 64.31212 | 60.02219 | 68.60206 |
| 20 | 48 | 168 | 2345 | 34.33848 | 29.28433 | 39.39263 |
| 21 | - | 180 | 3240 | 46.40197 | 40.06778 | 52.73617 |

图3-3-1 预测结果

由图3-3-1，可以得出当该市 X1=180（万人），人均月收入 X2=3240 的新的城市中销售其产品，求其销量的预测值为 46.40197万台，其置信度为 95%的置信区间为（40.06778 - 52.73617）万台。

姓名：
李金哲

学号：
201707010119

班级：
信 171

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。

四、
1)

表4-1-1是原有变量的相关系数矩阵。可以看到:存在部分的相关系数都较高，各变量呈较强的线性关系，够从中提取公共因子，适合进行因子分析

| | | 表4-1-1 相关性矩阵 | | | | | | | |
|-----|----|--------------|-------|-------|-------|-------|-------|-------|-------|
| | | 相关性矩阵 | | | | | | | |
| 相关性 | V2 | 1.000 | .334 | -.055 | -.061 | -.289 | .199 | .349 | .319 |
| | V3 | .334 | 1.000 | -.023 | .399 | -.156 | .711 | .414 | .835 |
| | V4 | -.055 | -.023 | 1.000 | .533 | .497 | .033 | -.139 | -.258 |
| | V5 | -.061 | .399 | .533 | 1.000 | .698 | .468 | -.171 | .313 |
| | V6 | -.289 | -.156 | .497 | .698 | 1.000 | .280 | -.208 | -.081 |
| | V7 | .199 | .711 | .033 | .468 | .280 | 1.000 | .417 | .702 |
| | V8 | .349 | .414 | -.139 | -.171 | -.208 | .417 | 1.000 | .399 |
| | V9 | .319 | .835 | -.258 | .313 | -.081 | .702 | .399 | 1.000 |
| | V9 | .319 | .835 | -.258 | .313 | -.081 | .702 | .399 | 1.000 |

表4-1-2 巴特利特球度检验

| KMO 和巴特利特检验 | | |
|--------------|------|---------|
| KMO 取样适切性量数。 | | .569 |
| 巴特利特球形度检验 | 近似卡方 | 142.983 |
| | 自由度 | 28 |
| | 显著性 | .000 |

由表4-1-2，可知巴特利特球度检验统计量的观测值为142.983，相应的概率P-值接近0。如果显著性水平a为0.05，由于概率P值小于显著性水平a，则应拒绝原假设，认为相关系数矩阵与单位阵有显著差异。同时，KMO值为0.569，根据 Kaiser给出的KMO度量标准可知原有变量适合进行因子分析。

2)

由表 4-2-1 可以得出，:第 1 个因子的方差贡献为 3.096，解释原有 8 个变量总方差的 38.7%(即 $3.096 \div 8 \times 100\%$)，累计方差贡献率为 38.7%;第 2 个因子的方差贡献为 2.367，解释原有 8 个变量总方差的 29.59%(即 $2.367 \div 8 \times 100\%$)，累计方差贡献率为 68.294%[即 $(3.096+2.367) \div 8 \times 100\%$]. 其余数据含义类似。在初始解中由于提取了 8 个因子，原有变量的总方差均被解释，累计方差贡献率为 100%。

图 4.1 所示显示了旋转后所有因子的载荷图，可以直观看出：X6、X9，即人均衣着商品支出和人均非商品支出比较靠近两个因子坐标轴。

表4-2-1 总方差解释

| | | 初始特征值 | | | 提取载荷平方和 | | | 旋转载荷平方和 ^a |
|----|-------|--------|---------|-------|---------|--------|-------|----------------------|
| 成分 | 总计 | 方差百分 | 比 | 累积 % | 总计 | 方差百分 | 比 | 累积 % |
| 1 | 3.096 | 38.704 | 38.704 | 3.096 | 38.704 | 38.704 | 3.070 | |
| 2 | 2.367 | 29.590 | 68.294 | 2.367 | 29.590 | 68.294 | 2.380 | |
| 3 | .920 | 11.500 | 79.794 | | | | | |
| 4 | .706 | 8.824 | 88.618 | | | | | |
| 5 | .498 | 6.231 | 94.848 | | | | | |
| 6 | .230 | 2.874 | 97.722 | | | | | |
| 7 | .131 | 1.635 | 99.357 | | | | | |
| 8 | .051 | .643 | 100.000 | | | | | |

提取方法：主成分分析法。

a. 如果各成分相关时，则无法添加载荷平方和以获取总方差。

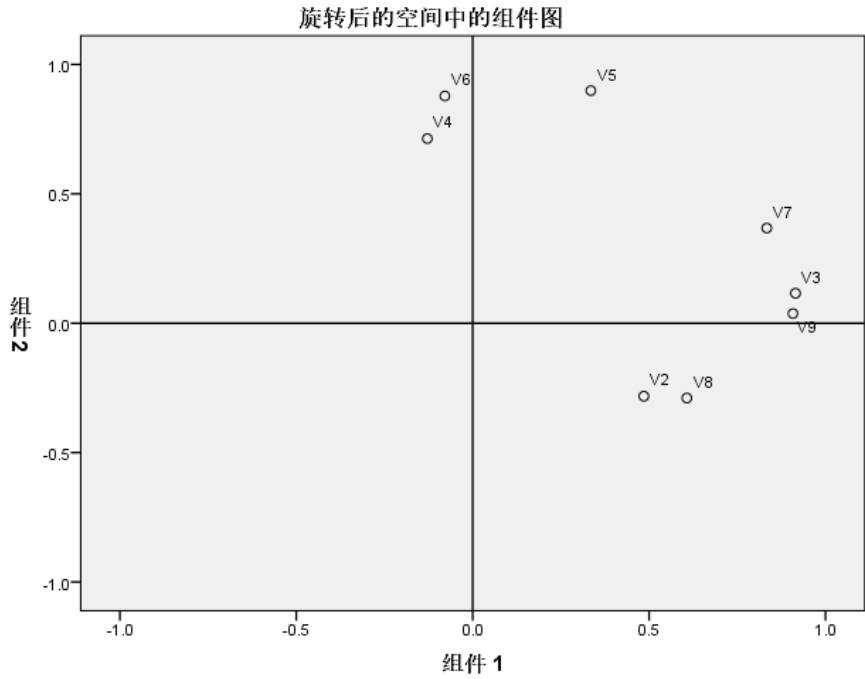


图4-2-1 旋转图

3)

根据谱系图4-3-1可得，上海为一类，广州为一类，其他的28个成员分为一类。总共为3类；具体的成员分类见表4-3-2。

姓名：
李金哲

学号：
201707010119

班级：
信 171

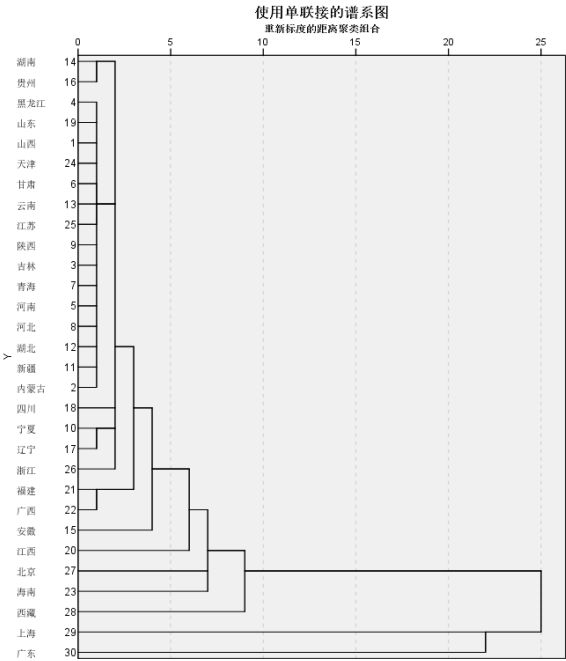


图 4-3-1 最短距离法的谱系聚类图

表4-3-1 聚类个案数

| 每个聚类中的个案数量 | | |
|------------|---|----|
| 聚类 | 1 | 28 |
| | 2 | 1 |
| | 3 | 1 |
| 有效 | | 30 |
| 缺失 | | 0 |

表4-3-1 聚类成员数据

聚类成员

| 个案 | 3 个聚类 | 2 个聚类 | 个案 | 3 个聚类 | 2 个聚类 | 个案 | 3 个聚类 | 2 个聚类 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1:山西 | 1 | 1 | 11:新疆 | 1 | 1 | 21:福建 | 1 | 1 |
| 2:内蒙古 | 1 | 1 | 12:湖北 | 1 | 1 | 22:广西 | 1 | 1 |
| 3:吉林 | 1 | 1 | 13:云南 | 1 | 1 | 23:海南 | 1 | 1 |
| 4:黑龙江 | 1 | 1 | 14:湖南 | 1 | 1 | 24:天津 | 1 | 1 |
| 5:河南 | 1 | 1 | 15:安徽 | 1 | 1 | 25:江苏 | 1 | 1 |
| 6:甘肃 | 1 | 1 | 16:贵州 | 1 | 1 | 26:浙江 | 1 | 1 |
| 7:青海 | 1 | 1 | 17:辽宁 | 1 | 1 | 27:北京 | 1 | 1 |
| 8:河北 | 1 | 1 | 18:四川 | 1 | 1 | 28:西藏 | 1 | 1 |
| 9:陕西 | 1 | 1 | 19:山东 | 1 | 1 | 29:上海 | 2 | 2 |
| 10:宁夏 | 1 | 1 | 20:江西 | 1 | 1 | 30:广东 | 3 | 2 |

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。

北京建筑大学 2019/ 2020 学年 第 2 学期 考试

数据分析 课程期末 答题纸

姓名：
李金哲

学号：
201707010119

班级：
信 171

- 1、考生必须将姓名、学号等项填写在装订线左侧内，不得超出，否则按违纪处理。
- 2、请将证件放在桌角处备查。
- 3、遵守考场规则。
- 4、凡违反考纪者按规定给予相应处分。