|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学号：** | **502025150026** | **姓名：** | **李振杉** | **专业：** | **管理科学与工程** |

第2次作业

**（本次作业提供所有数据的Excel文件；可以使用任意统计软件完成，如Excel、SPSS、R、Python、MATLAB等；将统计软件输出的关键图表截图粘贴进作业答案，并结合题目要求进行解释；如果用R或者Matlab做数据分析，请将code拷贝放入答案；图表的排版尽量美观）**

作业提交说明：1）直接在本Word文档中作答；

2）**10月30日晚20:00**前提交电子版（Word）；

3）用“学号+姓名+第1次作业”命名作业文件；

4）发送到邮箱a2970769529@163.com，同时抄送一份到邮箱252398366@qq.com；

1. （多元正态分布的性质）设3维随机变量*X*服从 ，其中；
2. 求的分布；
3. 与之间的关系；

以下是答案：

1. （多元样本的数字特征）从3维总体中随机抽取3个样本，样本数据矩阵是，求：
2. 样本均值；
3. 样本离差阵A；
4. 样本协方差阵S2；
5. 样本相关系数矩阵R；
6. 对样本数据做标准化处理，给出标准化处理后的数据矩阵、样本相关系数矩阵；

以下是答案：

3、（随机向量的均值与方差的性质）给定随机向量**X**，=，以及剖分=，其中，，的均值向量为，协方差矩阵为，且矩阵

，

1. 求，即的方差矩阵；
2. 求的均值方差矩阵，以及；

以下是答案：

* 1. （正态性评估与总体参数的假设检验）华盛顿每年春天举办樱花跑活动，全程10英里，2009年参与者的平均用时为103.5分钟；抽取2012年的150个参与者样本数据（表4），运用统计软件分析以下论断是否成立（先给出原假设和备择假设）：

(1) 跑步者的用时服从正态分布；

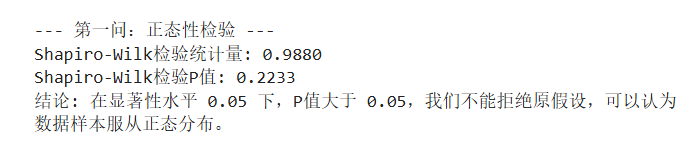
(2) 相较于2009年，2012年参与者平均用时更短；

表1 随机抽取2012年的150名参与者的跑步数据

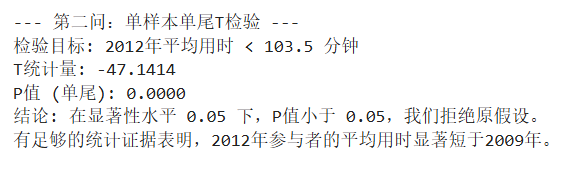
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 100.1 | 99.7 | 97.8 | 97.9 | 100.1 | 97.4 | 99.8 | 99.5 | 100.5 |
| 99.2 | 99.6 | 99.4 | 98.1 | 98.4 | 98.6 | 99.6 | 98.6 | 98.6 |
| 100.6 | 99.7 | 99.6 | 97.2 | 99.2 | 101.1 | 97.8 | 100.0 | 100.2 |
| 99.9 | 99.5 | 100.1 | 100.2 | 99.3 | 100.1 | 100.9 | 100.1 | 100.1 |
| 99.7 | 99.3 | 98.8 | 99.7 | 98.7 | 99.7 | 98.9 | 100.6 | 98.2 |
| 100.9 | 100.5 | 99.3 | 97.9 | 99.3 | 97.5 | 99.7 | 101.6 | 100.0 |
| 100.4 | 98.4 | 98.6 | 100.0 | 100.2 | 100.8 | 99.8 | 100.5 | 99.6 |
| 100.0 | 99.9 | 101.0 | 99.6 | 99.0 | 101.5 | 99.0 | 97.9 | 100.3 |
| 99.9 | 98.6 | 100.0 | 99.6 | 99.7 | 99.4 | 101.0 | 98.7 | 99.5 |
| 100.3 | 98.5 | 100.0 | 98.4 | 98.2 | 100.3 | 101.5 | 100.0 | 98.9 |
| 99.0 | 98.7 | 99.6 | 99.0 | 100.8 | 99.2 | 99.3 | 98.3 | 100.5 |
| 97.9 | 99.9 | 99.5 | 99.3 | 101.4 | 98.9 | 100.1 | 99.9 | 97.4 |
| 98.4 | 99.6 | 99.3 | 100.7 | 99.4 | 99.4 | 101.0 | 99.2 | 99.9 |
| 100.3 | 99.5 | 101.8 | 98.4 | 97.0 | 101.4 | 99.0 | 99.7 | 101.8 |
| 101.0 | 100.2 | 97.0 | 100.2 | 98.5 | 100.0 | 100.1 | 100.8 | 99.7 |
| 100.2 | 100.4 | 100.2 | 99.9 | 98.1 | 98.0 | 100.7 | 98.7 | 98.5 |
| 99.1 | 97.4 | 99.2 | 101.1 | 99.0 | 99.4 |  |  |  |

以下是答案：

1. ：



1. ：



* 1. （多元正态总体参数估计、均值和方差的假设检验）以2008年我国各省、市、自治区的城镇居民人均消费支出为例进行说明。反映城镇居民消费支出状况的指标主要有食品、衣着、居住、家庭设备用品及服务、医疗保健、交通和通信以及教育文化娱乐服务等八项消费支出指标，数据如表2所示。

1. 对上述八个指标的均值和协方差进行估计；（指出用到的估计量及其性质）
2. 对数据的正态性进行评估（，解释选择评估方法的原因）；
3. 按照区域对地区进行划分，分为四个总体（按照颜色依次分为北部地区、中东部地区、南部地区、西部地区；需要对上传数据增加所属区域标签）；选择符合正态分布的指标，比较四个总体的消费支出是否存在显著差异；如果存在显著差异，来自于哪些总体和哪些指标之间的差异？

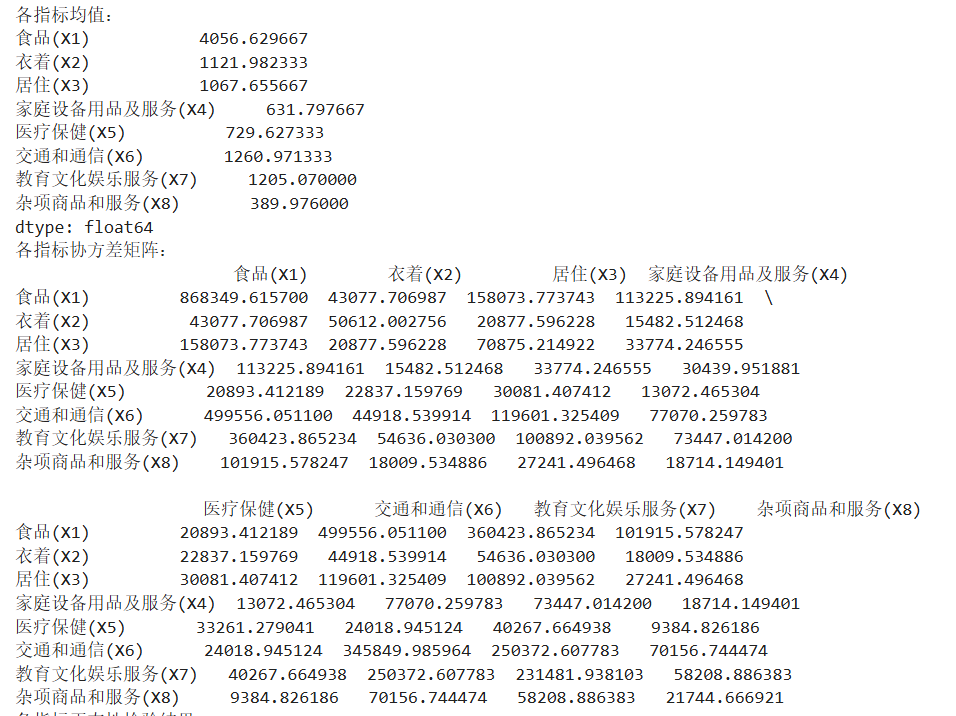
表2 2008年各地区城镇居民家庭平均每人全年消费性支出 单位：元

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 地区 | 食品 | 衣着 | 居住 | 家庭设备  用品及服务 | 医疗保健 | 交通和通信 | 教育文化  娱乐服务 | 杂项商品  和服务 |
| X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
| 北京 | 5561.54 | 1571.74 | 1286.32 | 1096.57 | 1563.10 | 2293.23 | 2383.52 | 704.24 |
| 天津 | 5005.09 | 1153.66 | 1528.28 | 817.18 | 1220.92 | 1567.87 | 1608.97 | 520.49 |
| 河北 | 3155.40 | 1137.22 | 1097.41 | 574.84 | 808.88 | 1062.31 | 946.38 | 304.28 |
| 山西 | 2974.76 | 1137.71 | 1250.87 | 471.65 | 769.79 | 931.33 | 1041.91 | 228.53 |
| 内蒙古 | 3553.48 | 1616.56 | 1028.19 | 672.64 | 869.71 | 1191.70 | 1383.53 | 512.81 |
| 辽宁 | 4378.14 | 1187.41 | 1270.95 | 507.40 | 913.13 | 1295.70 | 1145.46 | 533.29 |
| 吉林 | 3307.14 | 1259.62 | 1285.28 | 510.49 | 914.47 | 954.96 | 1071.80 | 425.30 |
| 黑龙江 | 3128.10 | 1217.04 | 941.25 | 494.49 | 864.89 | 749.05 | 906.19 | 321.95 |
| 上海 | 7108.62 | 1520.61 | 1646.19 | 1182.24 | 755.29 | 3373.19 | 2874.54 | 937.21 |
| 江苏 | 4544.64 | 1166.91 | 1042.10 | 813.45 | 794.63 | 1357.96 | 1799.75 | 458.10 |
| 浙江 | 5522.56 | 1546.46 | 1333.69 | 713.31 | 933.11 | 2392.63 | 2195.58 | 520.95 |
| 安徽 | 3905.05 | 1010.61 | 988.12 | 579.59 | 633.93 | 920.77 | 1160.14 | 325.82 |
| 福建 | 5078.85 | 1105.31 | 1300.10 | 722.17 | 540.63 | 1777.06 | 1453.18 | 523.83 |
| 江西 | 3633.05 | 969.58 | 851.15 | 623.17 | 483.96 | 872.57 | 945.99 | 337.91 |
| 山东 | 3699.42 | 1394.11 | 1247.04 | 806.35 | 799.79 | 1410.45 | 1277.43 | 372.01 |
| 河南 | 3079.82 | 1141.76 | 963.59 | 633.32 | 790.87 | 915.12 | 988.95 | 324.03 |
| 湖北 | 3996.27 | 1099.16 | 914.26 | 604.40 | 675.32 | 890.12 | 1037.24 | 260.74 |
| 湖南 | 3970.42 | 1090.72 | 960.82 | 674.84 | 790.95 | 971.05 | 1110.11 | 376.62 |
| 广东 | 5866.91 | 975.06 | 1748.16 | 947.54 | 836.39 | 2623.08 | 1936.38 | 594.45 |
| 广西 | 4082.99 | 772.28 | 891.33 | 603.84 | 529.36 | 1376.03 | 1081.54 | 290.04 |
| 海南 | 4226.90 | 491.84 | 1106.39 | 565.51 | 536.40 | 1303.50 | 930.87 | 247.08 |
| 重庆 | 4418.34 | 1294.30 | 1096.82 | 842.09 | 878.25 | 1044.36 | 1267.03 | 305.60 |
| 四川 | 4255.48 | 1042.45 | 819.28 | 590.51 | 564.93 | 1121.45 | 947.01 | 338.03 |
| 贵州 | 3597.94 | 851.50 | 836.54 | 525.70 | 471.39 | 871.15 | 934.73 | 260.27 |
| 云南 | 4272.29 | 1026.50 | 739.20 | 331.94 | 606.86 | 1216.46 | 732.95 | 150.42 |
| 西藏 | 4262.77 | 1011.82 | 634.94 | 310.22 | 317.08 | 966.74 | 419.59 | 400.38 |
| 陕西 | 3586.13 | 1047.61 | 1007.68 | 618.16 | 862.70 | 967.52 | 1281.58 | 400.68 |
| 甘肃 | 3183.79 | 1022.62 | 846.26 | 546.23 | 654.82 | 817.17 | 936.33 | 301.40 |
| 青海 | 3315.94 | 945.14 | 802.73 | 538.54 | 610.02 | 787.63 | 880.86 | 311.72 |
| 宁夏 | 3352.83 | 1178.88 | 1069.15 | 596.81 | 816.87 | 1096.32 | 1043.72 | 403.71 |
| 新疆 | 3235.77 | 1245.02 | 781.90 | 535.31 | 643.48 | 1003.89 | 812.36 | 411.63 |

资料来源：中国统计年鉴2009.

以下是答案：

1. ：



**所用估计量：**

样本均值向量。它是通过对样本中每个指标（每一列）的数据分别求算术平均数得到的向量。

**性质：**

无偏性：样本均值向量是总体均值向量 (μ) 的无偏估计，即 E(X̄) = μ。这意味着从长期来看，用大量不同样本计算出的均值向量的平均值会精确地等于总体的真实均值向量。

一致性：当样本量 n 增大时，样本均值向量会收敛于总体均值向量，是总体均值向量的一致估计。

**协方差矩阵的估计**

**所用估计量：**

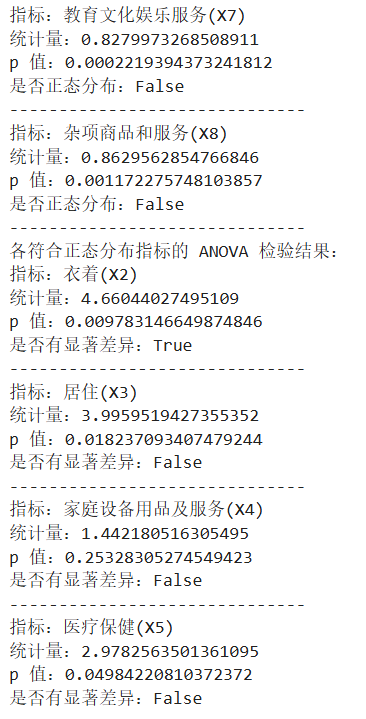
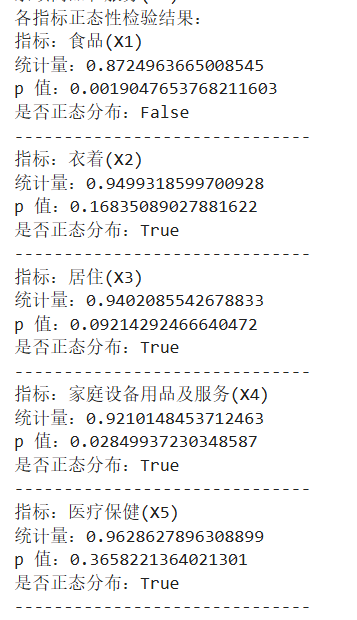
样本协方差矩阵 (S)。其对角线元素是每个指标的样本方差，非对角线元素是两两指标间的样本协方差。计算公式为 S = (1/(n-1)) \* A，其中 A 是样本离差阵。

**性质：**

无偏性：样本协方差矩阵 S 是总体协方差矩阵的无偏估计。分母采用 n-1 而不是 n 正是为了保证其无偏性。

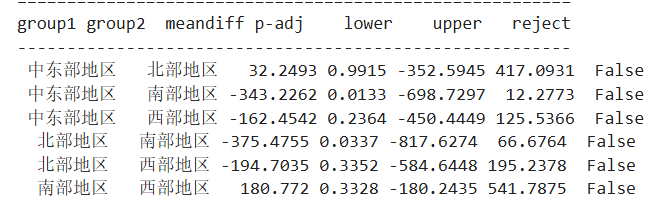
一致性：当样本量 n 增大时，样本协方差矩阵也会收敛于总体协方差矩阵。

1. :



选择 Shapiro - Wilk 检验是因为它在样本量较小时对正态性检验效果较好。从结果来看，只有 “衣着 (X2)”、“居住 (X3)”、“家庭设备用品及服务 (X4)” 和 “医疗保健 (X5)” 这几个指标满足正态分布。

1. ：



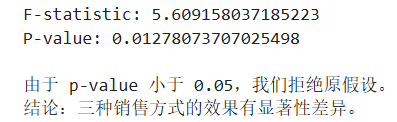
在0.01的显著性水平下，各地区两两之间均未达到显著差异（reject均为False）

* 1. （多元正态总体均值和方差的假设检验） 为了研究销售方式对商品销售额的影响，选择四种商品（甲、乙、丙和丁）按三种不同的销售方式（Ⅰ, Ⅱ和Ⅲ）进行销售。这四种商品的销售额分别为*x*1, *x*2, *x*3, *x*4，其数据见表3。试检验三种销售方式的效果有无显著性差异，对用于假设检验的统计量的进行解释。（）

表3 销售额数据

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编号 | 销售方式Ⅰ | | | | 销售方式Ⅱ | | | | | 销售方式Ⅲ | | | |
| *x*1 | *x*2 | *x*3 | *x*4 | *x*1 | | *x*2 | *x*3 | *x*4 | *x*1 | *x*2 | *x*3 | *x*4 |
| 1 | 125 | 60 | 338 | 210 | | 66 | 54 | 455 | 310 | 65 | 33 | 480 | 260 |
| 2 | 119 | 80 | 233 | 330 | | 82 | 45 | 403 | 210 | 100 | 34 | 468 | 295 |
| 3 | 63 | 51 | 260 | 203 | | 65 | 65 | 312 | 280 | 65 | 63 | 416 | 265 |
| 4 | 65 | 51 | 429 | 150 | | 40 | 51 | 477 | 280 | 117 | 48 | 468 | 250 |
| 5 | 130 | 65 | 403 | 205 | | 67 | 54 | 481 | 293 | 114 | 63 | 395 | 380 |
| 6 | 69 | 45 | 350 | 190 | | 38 | 50 | 468 | 210 | 55 | 30 | 546 | 235 |
| 7 | 46 | 60 | 585 | 200 | | 42 | 45 | 351 | 190 | 64 | 51 | 507 | 320 |

以下是答案：



7、（形象分析）作为例4.4.2爱情与婚姻问题研究的一部分，一个问卷调查的样本由新近结婚的30名男性和30名女性组成，请他们回答了下列问题：（1）要是全面地考虑，您如何描述自己对婚姻的“贡献”程度？（2）要是全面地考虑，您如何描述自己婚姻的“结果”？ 回答采用如下5级计分制：

1—很否定，2—中等程度否定，3—既不否定也不肯定，4—中等程度肯定，5—很肯定

同时请调查对象按下列5级计分制回答如下问题：（3）您感觉您伴侣的爱情“热度”如何？（4）您感觉您伴侣的爱情“可结伴”程度如何？

1—没有，2—很小，3—有些，4—较大，5—非常大

； ；

两个总体的定义：总体1—已婚男性；总体2—已婚女性；试做两总体的轮廓分析（取）：(由于打分取整数，数据不一定服从正态分布；样本容量较大，基于多元正态的方法可以近似采用)

(1) 分别画出丈夫和妻子的打分的轮廓图；

(2) “丈夫对妻子的打分”的轮廓是否平行、重合于“妻子对丈夫的打分”的轮廓；

(3) 如果两个轮廓是平行、重合的，进一步检验是否水平；

表4 婚姻调查数据

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 丈夫对妻子的打分 | | | | 妻子对丈夫的打分 | | | |
|  |  |  |  |  |  |  |  |
| 2 | 3 | 5 | 5 | 4 | 4 | 5 | 5 |
| 5 | 5 | 4 | 4 | 4 | 5 | 5 | 5 |
| 4 | 5 | 5 | 5 | 4 | 4 | 5 | 5 |
| 4 | 3 | 4 | 4 | 4 | 5 | 5 | 5 |
| 3 | 3 | 5 | 5 | 4 | 4 | 5 | 5 |
| 3 | 3 | 4 | 5 | 3 | 3 | 4 | 4 |
| 3 | 4 | 4 | 4 | 4 | 3 | 5 | 4 |
| 4 | 4 | 5 | 5 | 3 | 4 | 5 | 5 |
| 4 | 5 | 5 | 5 | 4 | 4 | 5 | 4 |
| 4 | 4 | 3 | 3 | 3 | 4 | 4 | 4 |
| 4 | 4 | 5 | 5 | 4 | 5 | 5 | 5 |
| 5 | 5 | 4 | 4 | 5 | 5 | 5 | 5 |
| 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| 4 | 3 | 5 | 5 | 4 | 4 | 4 | 4 |
| 4 | 4 | 5 | 5 | 4 | 4 | 5 | 5 |
| 3 | 3 | 4 | 5 | 3 | 4 | 4 | 4 |
| 4 | 5 | 4 | 4 | 5 | 5 | 5 | 5 |
| 5 | 5 | 5 | 5 | 4 | 5 | 4 | 4 |
| 5 | 5 | 4 | 4 | 3 | 4 | 4 | 4 |
| 4 | 4 | 4 | 4 | 5 | 3 | 4 | 4 |
| 4 | 4 | 4 | 4 | 5 | 3 | 4 | 4 |
| 4 | 4 | 4 | 4 | 4 | 5 | 4 | 4 |
| 3 | 4 | 5 | 5 | 2 | 5 | 5 | 5 |
| 5 | 3 | 5 | 5 | 3 | 4 | 5 | 5 |
| 5 | 5 | 3 | 3 | 4 | 3 | 5 | 5 |
| 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 |
| 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 |
| 3 | 3 | 5 | 5 | 3 | 4 | 4 | 4 |
| 4 | 4 | 3 | 3 | 4 | 4 | 5 | 4 |
| 4 | 4 | 5 | 5 | 4 | 4 | 5 | 5 |

以下是答案：

