

北京航空航天大学  
2019—2020 学年 第二学期期末

《应用统计学》  
考 试 A 卷

任课教师：王惠文、王珊珊

康雁飞、郝 壮

班 级 \_\_\_\_\_ 学 号 \_\_\_\_\_

姓 名 \_\_\_\_\_ 成 绩 \_\_\_\_\_

考试日期:2020 年 7 月 1 日

班号 \_\_\_\_\_ 学号 \_\_\_\_\_ 姓名 \_\_\_\_\_ 成绩 \_\_\_\_\_

## 《应用统计学》期末考试卷

注意事项：1、考试时间：14:00-17:00 共计 3 小时；

2、本次考试为线上开卷考试，考试过程中可使用统计分析软件（包含不限于 R、SPSS、Stata、Excel 等），考试结束后线上提交试卷。

一、（本题 20 分）设随机变量  $X$  的概率密度为  $f(x) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x < \theta \\ 0, & \text{else} \end{cases}$ ，其中未知

参数  $\theta > 0$ ,  $X_1, \dots, X_n$  是来自总体  $X$  的样本，求

（1） $\theta$  的矩估计，并回答其矩估计是否为相合的？

（2） $\theta$  的极大似然估计，并回答该极大似然估计是否无偏估计？是否为相合估计？

注：给出计算过程及依据，只给出结果没有成绩。

二、（本题 10 分）某一家公司生产袋装茶，现考察两部机器生产的袋装茶的工序质量，为此从两部机器生产的袋装茶中各自随机抽取 21 个样本，记录其重量（单位：克），如下所示：

机器 1: 3.45, 3.20, 3.22, 3.50, 2.95, 3.16, 3.20, 3.22, 2.98, 3.75, 3.38, 3.45, 3.48,

3.18, 3.90, 3.70, 3.28, 3.35, 3.20, 3.12, 3.25

机器 2: 3.22, 3.38, 3.30, 3.30, 3.34, 3.28, 3.30, 3.28, 3.19, 3.20, 3.29, 3.35, 3.16,

3. 34, 3. 35, 3. 30, 3. 05, 3. 33, 3. 27, 3. 28, 3. 25

假设两部机器生产的袋装茶重量分别服从正态分布  $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ ，且两部机器之间是独立的。请回答：

(1) 假设  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ，给出  $\mu_1 - \mu_2$  的置信度为  $1 - \alpha = 0.95$  的双侧置信区间；

(2) 生产工序方差是工序质量的一个重要度量，当方差较大时，需要对工序进行改进以减少方差，请构造两个总体方差之比  $\sigma_1^2 / \sigma_2^2$  的置信度为  $1 - \alpha = 0.95$  的双侧置信区间；

注：可利用统计软件进行计算，但需要给出过程及依据，只给出结果没有成绩。

三、（本题 10 分）某企业为比较两种方法对员工培训的效果，采用方法 1 对 15 名员工进行培训，采用方法 2 对 12 名员工进行培训，培训后的测试分数如下：

方法 1：56, 47, 42, 50, 47, 51, 52, 53, 42, 44, 45, 43, 52, 48, 44

方法 2：59, 52, 53, 54, 57, 56, 55, 64, 53, 65, 53, 57

假设两种方法培训效果分别服从正态分布  $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ ，且两种方法之间是独立的。在  $\alpha = 0.05$  的显著性水平下，

(1) 若假设  $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ，检验两种方法的培训效果  $\mu_1, \mu_2$  之间是否有显著差异；

(2) 采用两种培训方法，检验测试分数的方差  $\sigma_1^2, \sigma_2^2$  之间是否有显著差异。

注：可利用统计软件进行计算，但需要给出过程及依据，只给出结果没有成绩。

四、（本题 10 分）一家汽车企业的销售部门对东部地区、中部地区和西部地区的 400 个消费者作了抽样调查，得到如表 1 结果。

表 1 汽车调查数据

汽车价格	东部地区	中部地区	西部地区
10 万元以下	20	40	40
10 万~20 万元	50	60	50
20 万~30 万元	30	20	20
30 万元以上	40	20	10

在  $\alpha = 0.05$  的显著性水平下，检验地区与所购买汽车的价格是否独立。

注：可利用统计软件进行计算，但需要给出步骤过程及依据，只给出结果没有成绩。

五、（本题 50 分）在“十项全能成绩.csv（或：十项全能成绩.txt，十项全能成绩.sav）”数据集中，包含了某运动会中 34 名运动员的十项全能（100 米跑、跳远、铅球、跳高、400 米跑、110 米跨栏、铁饼、撑杆跳高、标枪、1500 米跑）成绩及总成绩数据。此外，还给出了每个运动员的综合评级，取值分别为 1 级或 2 级。上述 12 个变量的具体描述如表 2 所示。请结合以下分析思路，对该数据给出一份综合分析报告；并注意在报告中，重视运用描述性统计技术。

（1）运用主成分分析方法，将运动员的 10 项全能成绩数据从 10 维降至 2 维，从而绘制主平面图（二维主成分得分图），并对计算结果进行分析。

（2）采用聚类分析方法，将 34 位运动员按照成绩情况分成 5 类，并对各类运动员的特点进行分析。

（3）通过建立线性回归模型，分析影响运动员总成绩的主要因素。

（4）采用 Fisher 判别分析或者逻辑回归模型，建立运动员综合评级的判别模型，并对计算结果进行分析。

表 2 “十项全能成绩”数据集的各变量具体描述

编号	变量名称	变量描述	编号	变量名称	变量描述
1	r100 米	100 米跑成绩	7	铁饼	铁饼成绩
2	跳远	跳远成绩	8	撑杆跳	撑杆跳高成绩
3	铅球	铅球成绩	9	标枪	标枪成绩
4	跳高	跳高成绩	10	r1500 米	1500 米跑成绩
5	r400 米	400 米跑成绩	11	总成绩	十项全能总成绩
6	跨栏	110 米跨栏成绩	12	综合评级	运动员综合评级 取值为 1 或 2 级