

北京航空航天大学  
2021-2022 学年 第二学期期末

《应用统计学》  
考 试 A 卷

任课教师:

王惠文、秦中峰、王珊珊、康雁飞、郝 壮

班 级\_\_\_\_\_学 号 \_\_\_\_\_

姓 名\_\_\_\_\_成 绩 \_\_\_\_\_

考试日期:2022 年 6 月 23 日

## 《应用统计学》期末考试 A 卷

注意事项：

- 1、考试时间：10:20-12:20 共计 2 小时；
- 2、本次考试为线上开卷考试，可使用不带编程功能的科学计算器，可查看事先准备好的线下资料，但不能上网查询资料，电脑端仅限于查看试卷；
- 3、答题时可能会用到的下侧分位点数值为：

$$t_{0.975}(49) = 2.01, t_{0.975}(48) = 2.01, t_{0.95}(49) = 1.68$$

$$t_{0.975}(55) = 2.00, t_{0.95}(56) = 1.68, t_{0.975}(5) = 2.57$$

$$\chi_{0.025}^2(5) = 0.83, \chi_{0.975}^2(5) = 12.83, \chi_{0.95}^2(2) = 5.99$$

$$F_{0.95}(2,9) = 4.26, F_{0.975}(48,44) = 1.80, F_{0.975}(49,45) = 1.79$$

$$F_{0.025}(48,44) = 0.56, \quad F_{0.95}(48,44) = 1.64, \quad F_{0.05}(48,44) = 0.61$$

$$F_{0.025}(49,45) = 0.56, \quad F_{0.95}(49,45) = 1.63, \quad F_{0.05}(49,45) = 0.62$$

- 4、按照答题顺序写在 A4 答题纸上，考试结束后拍照或扫描粘贴在答题纸保存成 pdf 文件后线上提交答卷。

在正式答题之前，请同学们阅读“北京航空航天大学线上考试诚信承诺书”，并在在A4答题纸上书写以下文字：

本人已知悉并将遵守《北京航空航天大学线上考试诚信承诺书》相关内容。

一、（本题 10 分）设  $X_1, X_2, \dots, X_9$  是来自  $N(\mu, \sigma^2)$  的简单随机样本。令

$$\bar{X}_1 = \frac{1}{6} \sum_{i=1}^6 X_i, \quad \bar{X}_2 = \frac{1}{3} \sum_{i=7}^9 X_i,$$

$$S_1^2 = \frac{1}{5} \sum_{i=1}^6 (X_i - \bar{X}_1)^2, \quad S_2^2 = \frac{1}{2} \sum_{i=7}^9 (X_i - \bar{X}_2)^2, \quad Z = \frac{\sqrt{2}(\bar{X}_1 - \bar{X}_2)}{S_1}.$$

请回答：(1)  $\bar{X}_1 - \bar{X}_2$  的分布；(2)  $Z$  的分布；(3)  $S_1^2/S_2^2$  的分布。

注：请给出过程和结果，只给出结果者会扣分

二、（本题 20 分）设随机变量  $X$  的概率密度为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta}-1}, & 0 < x < 1, 0 < \theta < \infty \\ 0, & \text{其他.} \end{cases}$$

$X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本。

- （1）求出参数  $\theta$  的矩估计量  $\hat{\theta}_1$ ，并回答：该矩估计是否为相合估计？
- （2）求出参数  $\theta$  的极大似然估计  $\hat{\theta}_2$ ，并回答：该极大似然估计是否无偏估计？
- （3）求极大似然估计  $\hat{\theta}_2$  的方差，并据此判断  $\hat{\theta}_2$  的渐近分布。

提示：求解过程中可能利用到以下极限：

$$\lim_{x \rightarrow 0} x^{\frac{1}{\theta}} \ln(x) = 0, \lim_{x \rightarrow 0} x^{\frac{1}{\theta}} [\ln(x)]^2 = 0.$$

注：给出计算过程及判断依据，只给出结果者会扣分。

三、（本题 10 分）假设某同学在进行某种液体的沸腾温度实验时，进行 6 次实验观测到的读数（以摄氏度为单位）为：

102.5、101.7、103.1、100.9、100.5、102.2.

假设实验得到该液体的沸腾温度读数服从正态分布  $N(\mu, \sigma^2)$ ，那么在  $1 - \alpha = 95\%$  置信水平下：

- （1）总体均值  $\mu$  的区间估计是多少？
- （2）总体方差  $\sigma^2$  的区间估计是多少？
- （3）若要求在 95% 置信水平下，总体均值  $\mu$  区间估计的长度不超过 1 摄氏度，该学生至少需要进行多少次实验？

注：可利用计算器进行计算，但需要给出过程及依据，只给出结果会扣分。

四、(本题 10 分) 为了评估①吸烟的健康危害信息和②戒烟方法信息对烟民对某种戒烟产品支付意愿的影响, 进行如下实验:

从我国烟民中简单随机抽取三组人进入 1 组控制组和 2 组实验。为控制组提供安慰剂信息; 对实验组 1 提供吸烟的健康危害信息; 对实验组 2 提供戒烟方法信息。测量所有人对一个疗程戒烟药品的支付意愿 (元), 得到如表 1 统计信息

表 1. 统计信息

组别	观测数	样本均值 (元)	样本标准差
控制组	56	1287.946	717.6428
实验组 1	49	1192.602	720.315
实验组 2	45	1470.833	570.5858

假设各组烟民对戒烟产品的支付意愿服从正态分布, 即控制组、实验组 1、实验组 2 这三组烟民对该戒烟产品的支付意愿分别服从  $N(\mu_1, \sigma_1^2)$ ,  $N(\mu_2, \sigma_2^2)$ ,  $N(\mu_3, \sigma_3^2)$ , 且各组人的支付意愿是独立的。在  $\alpha = 0.05$  的显著性水平下,

(1) 能否拒绝原假设  $H_0: \mu_1 = 1300$  vs.  $H_1: \mu_1 \neq 1300$ ? 分别给出检验统计量的观测值和拒绝域;

(2) 能否拒绝两个实验组的总体方差相等这一原假设? 即  $H_0: \sigma_2^2 = \sigma_3^2$  vs.  $H_1: \sigma_2^2 \neq \sigma_3^2$ ?

注: 可利用计算器进行计算, 但需要给出过程及依据, 只给出结果没有成绩。

五、(本题 12 分) 设下面是 5 个样本之间的距离矩阵。

$$(d_{ij})_{5 \times 5} = \begin{pmatrix} 0 & 9 & 3 & 6 & 11 \\ & 0 & 7 & 5 & 10 \\ & & 0 & 9 & 2 \\ & & & 0 & 8 \\ & & & & 0 \end{pmatrix}$$

请利用最短距离法, 进行系统聚类, 并绘制谱系聚类图。

六、(本题 12 分) 某网站根据性能、外观等特性对不同的超薄笔记本电脑进行评估。每种特性的评估分采用百分制，对于每个型号的笔记本电脑将得到一个总体评估分。现收集 10 台不同型号的超薄笔记本电脑的性能、外观和网站评估分的数据。以性能评估分、外观等级评估分为自变量，某网站评估分为因变量，建立回归模型。利用统计软件得到表 2 的分析结果：

表 2. 统计软件回归分析结果

SUMMARY OUTPUT						
回归统计						
Multiple R	0.915					
R Square	0.837					
Adjusted R <sup>2</sup>	0.790					
标准误差	1.673					
观测值	10					
方差分析						
	df	SS	MS	F	Significance F	
回归分析	A1	100.51	A2	A3	0.0018	
残差	7	A4	A5			
总计	9	120.10				
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	39.98	7.86	5.09	0.00	21.41	58.56
性能评估分	0.11	0.04	B1	0.03	0.02	0.20
外观等级评估分	0.38	B2	3.49	0.01	0.12	0.64

根据表 2 的结果，回答如下问题：

- (1) 令变量  $Y$  = 网站评估分、变量  $X_1$  = 性能评估分、变量  $X_2$  = 外观等级评估分，写出估计的回归方程（即回归模型），并解释回归系数的含义。
- (2) 根据已有信息，计算出表 2 中 **A1**、**A2**、**A3**、**A4**、**A5**、**B1** 和 **B2** 的值；
- (3) F 检验的原假设是什么？请对该回归模型进行 F 检验，并解释检验结果的含义。(显著性水平为 0.05)
- (4) 检验变量  $X_1$  和  $X_2$  的回归系数的显著性，并解释检验结果的含义。(显著性水平为 0.05)
- (5) 当一台新的笔记本性能评估分为 70 分 ( $X_1 = 70$ )，外观等级评估分为 90 分 ( $X_2 = 90$ ) 时，期望的网站平均评估分是多少？

七、(本题 14 分) 某位分析人员收集了 2020 年 31 个地区的 5 项经济指标变量, 分别是: 人均地区生产总值 (单位: 元)、财政收入 (单位: 万元)、年末总人口 (单位: 万人)、居民人均消费支出 (单位: 元/人)、社会消费品零售总额 (单位: 亿元)。经过数据标准化和主成分分析, 得到这些变量协方差矩阵的特征值分别是:

$$\lambda_1 = 3.316, \lambda_2 = 1.519, \lambda_3 = 0.085, \lambda_4 = 0.068, \lambda_5 = 0.012$$

对应于 $\lambda_1$ 和 $\lambda_2$ 的特征向量请参见表 3。

表 3. 特征向量

$u_1$	0.080	0.482	0.536	0.088	0.518
$u_2$	0.791	0.346	-0.078	0.787	0.246

此外, 部分原始变量 $x_j$ 与第一主成分 $y_1$ 、第二主成分 $y_2$ 的相关系数请参见表 4。

表 4. 主成分与原始变量的相关系数

原始变量的编号 j	$r(x_j, y_1)$	$r(x_j, y_2)$
1	0.146	
2		0.426
3	0.976	-0.096
4	0.16	0.97
5	0.944	0.303

(1) 请在表 4 中, 填写  $r(x_1, y_2)$  和  $r(x_2, y_1)$  的数据。

(2) 请写出第一主成分 ( $F_1$ ) 和第二主成分 ( $F_2$ ) 关于原始变量的函数表达式, 并为第一主成分和第二主成分命名。

(3) 请计算第一主成分  $F_1$  的均值和方差, 以及前两个主成分的累计贡献率。

(4) 如果样本点 A 是在主平面图第一象限的右上方, 而样本点 B 是在主平面图第三象限的左下方, 请分别说明这两个地区的经济发展特征。

八、(本题 12 分) 生活在某岛屿中的企鹅有 3 种, 阿德利企鹅( $G_1$ )、纹颊企鹅( $G_2$ )、巴布亚企鹅( $G_3$ )。在这 3 个种类的企鹅中, 分别抽取样本容量为  $n_1 = 117$ ,  $n_2 = 55$ ,  $n_3 = 96$  的样本。为了进行判别分析, 选取了企鹅的四个尺寸测量指标为依据:  $X_1$ —喙长(毫米)、 $X_2$ —喙深(毫米)、 $X_3$ —鳍长(毫米)、 $X_4$ —体重(克)。设这三个总体的方差没有显著差别。采取 Fisher 判别法(又称: 线性判别分析方法)计算, 部分结果见表 5 和表 6。

表 5. 未标准化的典型判别函数系数

	函数	
	1	2
$X_1$	0.100	0.446
$X_2$	-1.049	-0.002
$X_3$	0.079	-0.029
$X_4$	0.001	-0.002
(常量)	-7.707	-6.955

表 6. 组重心处的判别函数值

企鹅种类	函数	
	1	2
$G_1$	-3.249	-1.216
$G_2$	-1.900	3.088
$G_3$	5.048	-0.288

- (1) 请简述 Fisher 判别法与主成分分析的原理和区别。
- (2) 请写出第一典型判别函数和第二典型判别函数的数学表达式。
- (3) 某只企鹅的尺寸测量数据为:  $X_1 = 42$ ,  $X_2 = 19.5$ ,  $X_3 = 200$ ,  $X_4 = 4050$ 。假设 3 种类型的企鹅在第一判别函数上的方差没有显著差异, 请采用第一典型判别函数, 判断该企鹅属于哪一类型?