

统计与数据科学学院本科生 23——24 学年第 1 学期数据采集方法课程期末考试试卷（A 卷）

专业:

年级:

学号:

姓名:

成绩:

草稿区

得分

一 、(8 分) 一位大夫从去过他们小儿门诊所就诊的 2 到 6 岁的儿童里，简单随机抽取了 240 个。发现这些孩子当中开始走路的频数分布如下，计算可得小孩独立走路的样本均值为 12.08，样本方差为 3.71。假如有一位研究者打算在另一个地方展开此类研究,从幼儿保健机构登记的两岁到六岁孩子的名单中，利用简单随机抽样方式抽取样本，估计孩子开始走路的年龄，希望绝对误差不超过 0.2，请利用上述数据,估算需要多大的样本量？

年龄(月)	9	10	11	12	13	14	15	16	17	18	19	20
孩子的数目	13	35	44	69	36	24	7	3	2	5	1	1

得分

二 、(14 分) 在一项关于某地区居民患有某一类关节炎的研究中，依据经验把总体分为两层。假如第一层发病率高，包含 20000 个居民，第二层发病率低，包含 150000 个居民。现在每层简单随机抽取 200 个单元，发现第一层发病率为 0.1，第二层为 0.02。

- (1) 请依据此数据，估计该地该类关节炎平均发病率；
- (2) 若打算进行更细致的调查，总样本量为 1000，使用奈曼分配将样本量分到两个层，如何分配？

得分

三、（12 分）某单位欲征求员工对一项改革措施的意见,该单位有 12 个下属二级单位,各二级单位人数见下表,单位员工总数为 800 人。现采用有放回不等概率抽样从 12 个二级单位中抽取 4 个,每个二级单位被抽中的概率与其人数成比例。然后在被抽中的每个二级单位中,采用简单随机抽样抽取 10 人,征询对该项改革措施意见,调查数据见下表。试根据调查结果估计该单位赞成此项改革措施的人员总数并给出估计量的方差估计。

车间	1	2	3	4	5	6	7	8	9	10	11	12
人数	45	82	52	55	62	96	85	73	50	76	64	60
被抽中	否	否	是	否	否	是	否	否	否	是	是	否
赞成人数			4			3				8	6	

得分

四、（12 分）现有两个 2^{6-2} 设计 A 和 B，其定义关系分别为 A：5=1234，6=124 B：5=123，6=124。请写出这两个设计的定义对照子群、字长型、分辨度，并在 MA 准则下比较这两个设计的好坏。

得 分

五、（8 分）阅读最后一页调查问卷，调查者希望了解消费者在机票盲盒购买过程中的关注点以及购买影响因素，请指出问卷存在的问题和不足。

得分

六、（15 分）一位营养学家在某次试验中考虑三个因子对于面团粘性（viscosity）的影响，因子的名称及水平在下表给出。他采用一个完全因析设计安排试验，结果如下。（1）请计算因子 C 的主效应以及交互效应 B×C；（2）做出因子 B 对 C 的交互效应图；（3）若已知每个观测的标准差是 1.5，计算（1）中估计量的标准误差。

Factor	Level	
	—	+
A. Temperature	25	35
B. Mixing speed	Low	High
C. Time (sec.)	60	90

Run	Factor			Viscosity
	A	B	C	
1	—	—	—	60
2	+	—	—	62
3	—	+	—	82
4	+	+	—	86
5	—	—	+	85
6	+	—	+	85
7	—	+	+	61
8	+	+	+	61

得 分

七、（15 分）某会计事务所对所报名的 30 名审计员准备同时比较 3 种培训方法的效果，他们按随机化完全区组设计的要求做如下安排：

把 30 名审计员按毕业年限从小到大排队；均分为 10 组，毕业年限最短的 3 人分到第一个区组，而毕业年限最长的 3 人分到第十个区组；进而对把每个区组内的 3 名审计员随机地安排到三个方法中去。在培训结束时，每名审计员都要去分析一个复杂的案例，根据分析结果，评分小组给每名审计员评分，结果如下表所示。注意表中倒数第二列为相应行总量，最后一列为相应行均值，而倒数第二行为相应列总量，最后一行为相应列均值。

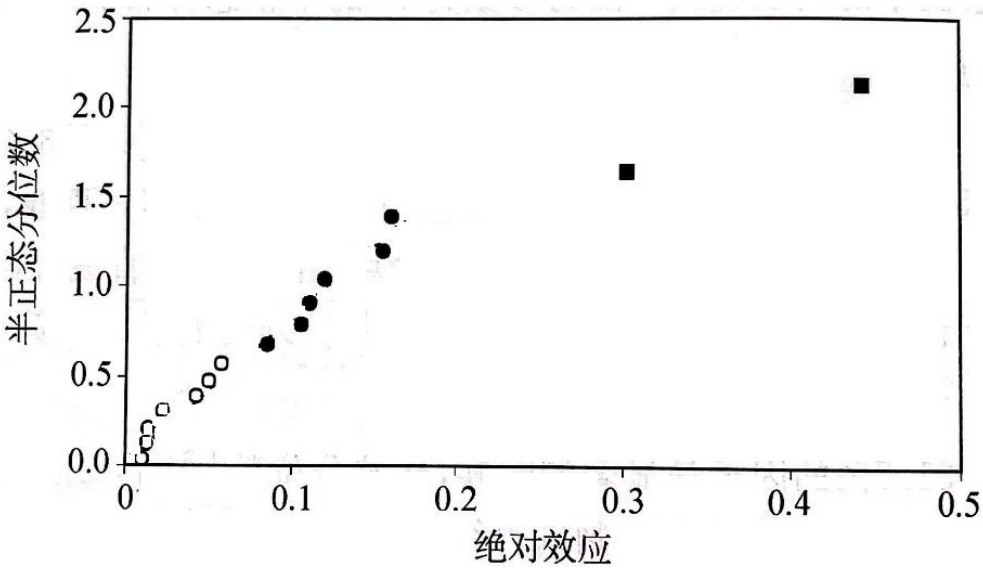
- (1) 你为什么相信“按毕业年限”划分区组是合理的？
- (2) 写出该设计的统计模型。
- (3) 求出第一个处理效应和区组效应的估计值。

	1	2	3	4	5	6	7	8	9	10	$y_{i.}$	$\bar{y}_{i.}$
1	73	76	75	74	76	73	68	64	65	62	706	70.6
2	81	78	76	77	71	75	72	74	73	69	746	74.6
3	92	89	87	90	88	86	88	82	81	79	861	86.1
$y_{.j}$	246	243	238	241	235	234	228	220	219	209	$y_{..} = 2313$	
$\bar{y}_{.j}$	82	81	79.3	80.3	78.3	78	76	73.3	73	69.6	$\bar{y}_{..} = 77.1$	

八、（8 分）在考虑因子效应的显著性时，可以利用图方法，比如正态概率图和半正态概率图。在一个实际问题中，采用 16 行 15 列的饱和正交表安排试验考察 15 个因子的效应，下表是各因子效应的估计值，下图为其半正态概率图，请根据此图判断哪些效应可能是显著的？并给出相应的原理解释。

得分

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
估计值	-0.11	0.01	0.16	-0.12	0.44	0.11	0.30	-0.09	-0.02	-0.16	0.06	0.05	0.10	0.03	0.06



得分

九、（8 分）林务员想要估计在某区域内树木的平均年龄。因为只能通过数年轮才能确知树龄，这种方法较难以大量实施，然而一般树龄越大直径越长，并且直径易于测量。林务员测量了所有 1132 棵树的直径，并算得直径的总体平均值为 10.3，然后他们随机抽取 20 棵树利用年轮来测量树龄。对于这片区域内树木的平均年龄，请给出两种不同的估计方法。数据中相应的样本均值以及样本方差、样本协方差分别为

$\bar{x} = 9.405; \quad \bar{y} = 107.4; \quad s_x^2 = 3.345; \quad s_y^2 = 821.52; \quad s_{xy}^2 = 40.972$

编号	直径 x	树龄 y	编号	直径 x	树龄 y
1	12.0	125	11	5.7	61
2	11.4	119	12	8.0	80
3	7.9	83	13	10.3	114
4	9.0	85	14	12.0	147
5	10.5	99	15	9.2	122
6	7.9	117	16	8.5	106
7	7.3	69	17	7.0	82
8	10.2	133	18	10.7	88
9	11.7	154	19	9.3	97
10	11.3	168	20	8.2	99

调查问卷

1. 您的性别为？ [单选题] *
☐男 ☐女
2. 请问您的年龄？ [单选题] *
☐20 岁以下 ☐20 岁-30 岁 ☐30 岁-40 岁 ☐40 岁以上
3. 请问您的身份是？ [单选题] *
☐学生 ☐上班族 ☐自营户 ☐其他
4. 您是否买过盲盒？ [单选题] *
☐买过 ☐没买过
5. 您了解过哪些类型的盲盒？ [多选题] *
☐手办 ☐图书 ☐飞机票 ☐美妆 ☐餐饮 ☐线上 app 内（类似游戏抽卡） ☐其他
6. 您购买过几次盲盒？ [单选题] *
☐0 次 ☐1-5 次 ☐5-10 次 ☐10 次以上
7. 您为盲盒花费了？ [单选题] *
☐0 元 ☐300 元以内 ☐300—1000 元 ☐1000 元以上
8. 您购买飞机票盲盒后是否愿意前往？ [单选题] *
☐愿意 ☐不愿意
9. 影响您前往的因素是？ [多选题] *
☐目的地 ☐时间 ☐近期消费能力 ☐其他
10. 您希望飞机票盲盒的出游时间是？ [多选题] *
☐工作日 ☐周末 ☐十一假期 ☐春节假期 ☐其他法定节假日
11. 您愿意购买飞机票盲盒旅游的目的地是？ [多选题] *
☐缓解工作压力 ☐陪家人出游 ☐打发闲散时间 ☐其他
12. 您对飞机票盲盒这种形式是否满意？ [单选题] *
☐满意 ☐不满意
13. 机票盲盒吸引您的地方是？ [多选题] *
☐价格便宜 ☐新鲜感 ☐其他