

中国科学技术大学

2015—2016学年第二学期考试试卷

考试科目 _____ 抽样调查 _____ 得分 _____

所在系 _____ 姓名 _____ 学号 _____

考试时间: 2016年4月27日上午9:45—12:15; 使用简单计算器

一. (20分) 为了调查某班级本科生的平均身高, 从该班级 $N = 100$ 人中随机抽取了10个样本, 记录如下 (单位cm): 170, 160, 172, 163, 162, 183, 175, 163, 165, 187。

(1) 计算学生样本平均身高 \bar{y} ;

(2) 假如总体身高的标准差为10, 试计算 \bar{y} 的方差, 并构造总体均值的置信度为95%的置信区间;

(3) 假如总体身高的标准差未知, 试计算 \bar{y} 的方差, 并构造总体均值的置信度为95%的置信区间。

二. (20分) 如上题, 若该班级学生的平均体重为60kg, 且被抽取学生相对应的体重为: 59, 55, 62, 57, 56, 70, 61, 57, 58, 75。

(1) 计算该班级学生平均身高的回归估计 \bar{y}_{lr} ;

(2) 试计算 \bar{y}_{lr} 的方差, 并利用回归估计构造总体均值的置信度为95%的置信区间;

(3) 比较回归估计 \bar{y}_{lr} 与简单估计 \bar{y} 的效率。

三. (16分) 调查某小区居民关于汽车消费的支出, 以居民户为抽样单元, 根据经济及收入水平将居民划为4层, 每层按简单随机抽样抽取5户, 调查结果如下表格,

h	N_h	y_{hi}				
		1	2	3	4	5
1	50	10	20	0	15	5
2	100	20	50	60	10	10
3	150	15	0	5	30	20
4	200	25	10	20	25	30

其中 h 代表层, N_h 为每层的居民户数, y_{hi} 为每户汽车消费支出 (单位: 万元)。

(1) 请估计该地区居民的平均汽车消费支出及其95%的置信区间;

(2) 试计算该分层抽样的设计效应。

四. (24分) 如上题(题目三), 若样本量仍为20, 对各层的抽样平均每户费用比为1:2:3:4, 试求:

- (1) 试求各层按比例分配和按最优分配时的样本量;
- (2) 若要求在95%置信度下, 相对误差不超过10%, 则层按比例分配和按奈曼分配时, 总样本量分别为多少?

五. (20分) 设总体中有 N 个单元, 第 i 个单元的规模是 Z_i , 某指标的测量值为 Y_i 。现考虑耶茨-格伦迪逐个抽样法:

- (1) 以概率 Z_i 从总体中抽取第一个单元, 编号记为1;
- (2) 以概率 $Z_i/(1 - Z_1)$ 从余下的 $N - 1$ 个单元中抽取第二个单元, 编号记为2;
- (3) 以概率 $Z_i/(1 - Z_1 - Z_2)$ 从余下的 $N - 2$ 个单元中抽取第三个单元, 编号记为3;
-
- (n) 以概率 $Z_i/(1 - \sum_{k=1}^{n-1} Z_k)$ 从余下的 $N - (n - 1)$ 个单元中抽取第 n 个单元, 编号记为 n .

记取出的顺序样本为: y_1, y_2, \dots, y_n , 对应的 z 值为: z_1, z_2, \dots, z_n 。令

$$\begin{aligned} t_1 &= y_1/z_1 \\ t_2 &= y_1 + y_2/[z_2/(1 - z_1)] \\ t_3 &= y_1 + y_2 + y_3/[z_3/(1 - z_1 - z_2)] \\ &\dots \dots \\ t_n &= \sum_{k=1}^{n-1} y_k + y_n/[z_n/(1 - \sum_{k=1}^{n-1} z_k)]. \end{aligned}$$

构造统计量

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n t_i,$$

那么

- (1) 试证明 \hat{Y} 是总体总量 Y 的一个无偏估计;
- (2) 试计算 \hat{Y} 的方差;
- (3) 试证明

$$A = \frac{1}{n(n-1)} \sum_{i=1}^n (t_i - \hat{Y})^2$$

是 \hat{Y} 方差的一个无偏估计。

附录 分位数: $u_{0.025} = 1.960$, $u_{0.05} = 1.645$, $t_{0.025}(10) = 2.228$, $t_{0.05}(10) = 1.812$, $t_{0.025}(11) = 2.201$, $t_{0.05}(11) = 1.796$.