

数据采集方法作业

姓名：蒋贵豪 学号：B+X9bo

2021 年 10 月 18 日

题目 1. 某市进行家庭收入调查，分城镇居民和农村居民两部分抽样，在全部城镇 $N_1 = 23560$ 户中抽取，在全部农村 $N_2 = 148420$ 户中抽取（均按简单随机抽样进行），若设城镇居民抽取一户的平均费用是 $c_1 = 1$ ，农村抽取一户的平均费用为 $c_2 = 2$ 。城镇居民标准差 $S_1 = 3000$ ，农村居民标准差 $S_2 = 2500$ 。若要求的 95% 绝对误差限 $d = 200$ ，求比例分配、Neyman 最优分配及一般最优分配下所需要的总样本量，又计算各种分配形式下的总费用（设 $c_0 = 0$ ）。

解答. 根据已知的数据，通过计算可以得到下表中的结果：

h	N_h	W_h	S_h	c_h	$W_h S_h$	$W_h S_h / \sqrt{c_h}$	$W_h S_h \sqrt{c_h}$
1	23560	0.137	3000	1	411.0	411.0	411.0
2	148420	0.863	2500	2	2157.5	1525.6	3051.2
求和	171980	1			2568.5	1936.6	3462.2

(1) 比例分配的总样本量为：

$$n = \frac{\sum_{h=1}^L W_h S_h^2}{\left(\frac{d}{z_{\frac{\alpha}{2}}}\right)^2 + \frac{1}{N} \sum_{h=1}^L W_h S_h^2} = \frac{0.137 \times 3000^2 + 0.863 \times 2500^2}{\left(\frac{200}{1.96}\right)^2 + \frac{1}{171980} \times (0.137 \times 3000^2 + 0.863 \times 2500^2)} \approx 634$$

而 $w_h = W_h$ ，于是比例分配的总费用为

$$F = \sum_{h=1}^L [n W_h] c_h = [634 \times 0.137] \times 1 + [634 \times 0.863] \times 2 = 1181$$

(2)Neyman 分配的总样本量为:

$$n = \frac{(\sum_{h=1}^L W_h S_h)^2}{(\frac{d}{z_{\frac{\alpha}{2}}})^2 + \frac{1}{N} \sum_{h=1}^L W_h S_h^2} = \frac{(0.137 \times 3000 + 0.863 \times 2500)^2}{(\frac{200}{1.96})^2 + \frac{1}{171980} \times (0.137 \times 3000^2 + 0.863 \times 2500^2)} \approx 631$$

又由:

$$w_h = \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}$$

于是, Neyman 分配的总费用为:

$$F = \sum_{h=1}^L [n \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}] c_h = [634 \times \frac{411}{2568.5}] \times 1 + [634 \times \frac{2157.5}{2568.5}] \times 2 = 1161$$

(3) 一般最优分配的总样本量为:

$$n = \frac{(\sum_{h=1}^L W_h S_h \sqrt{c_h})(\sum_{h=1}^L W_h S_h / \sqrt{c_h})}{(\frac{d}{z_{\frac{\alpha}{2}}})^2 + \frac{1}{N} \sum_{h=1}^L W_h S_h^2} = \frac{(411 + 3051.2) \times (411 + 1525.6)}{(\frac{200}{1.96})^2 + \frac{1}{171980} \times (0.137 \times 3000^2 + 0.863 \times 2500^2)} \approx 642$$

又由:

$$w_h = \frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^L W_h S_h / \sqrt{c_h}}$$

于是, 一般最优分配的总费用为:

$$F = \sum_{h=1}^L [n \frac{W_h S_h / \sqrt{c_h}}{\sum_{h=1}^L W_h S_h / \sqrt{c_h}}] c_h = [642 \times \frac{411}{1936.6}] \times 1 + [642 \times \frac{1525.6}{1936.6}] \times 2 = 1148$$

题目 2. 调查某地区居民牛奶消费支出，以居民户为抽样单元，分为 4 层，每层抽 10 户。数据如下图所示；如果在得到第一次调查结果后，想再做一次调查，并且样本量仍为 40 个，则按比例分配和 Neyman 分配，各层的样本量应为多少？

层	居民户总数	样本户牛奶年消费支出									
		1	2	3	4	5	6	7	8	9	10
1	200	10	40	0	110	15	10	40	80	90	0
2	400	50	130	60	80	100	55	160	85	160	170
3	750	180	260	110	0	140	60	200	180	300	220
4	1500	50	35	15	0	20	30	25	10	30	25

解答. 根据已知的数据，通过计算可以得到下表中的结果：

h	N_h	W_h	S_h	$W_h S_h$
1	200	0.07018	40.3078	2.8288
2	400	0.14035	46.5475	6.5329
3	750	0.26316	90.5845	23.8382
4	1500	0.52632	13.9044	7.3182
求和	2850	1		40.5181

(1) 比例分配的各层样本量为： $n_h = [nW_h]$ ，通过表格中的数据： $n_1 = [40W_1] = 3$ ， $n_2 = [40W_2] = 6$ ， $n_3 = [40W_3] = 10$ ， $n_4 = [40W_4] = 21$ 。

(2) Neyman 分配的各层样本量为：

$$w_h = \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}$$

于是， $n_1 = [40w_1] = 3$ ， $n_2 = [40w_2] = 6$ ， $n_3 = [40w_3] = 24$ ， $n_4 = [40w_4] = 7$ 。

题目 3. 某公司进行财务审计，需要对原始凭证进行审核。该公司先后有两名出纳，由 A 出纳登记的原始凭证占 70%，B 出纳登记的原始凭证占 30%。审计人员从原始凭证中随机抽出 100 份，结果发现，由 A，B 出纳登记的原始凭证分别为 43 份和 57 份，差错分别为 1 份和 2 份。

- (1) 用简单随机抽样的公式估计登记原始凭证的差错率，并计算估计的标准差；
- (2) 用事后分层的公式估计登记原始凭证的差错率，并计算估计的标准差。

有限总体校正系数 $1 - f \approx 1$

解答. (1) 简单随机抽样估计登记原始凭证的差错率为：

$$p = \frac{1 + 2}{100} = 3\%$$

方差为：

$$V(p) = \frac{1 - f}{n - 1} p(1 - p) = \frac{0.03 \times 0.97}{99} \approx 2.94 \times 10^{-4}$$

标准差为： $\sqrt{V(p)} \approx 0.0171$

(2) 事后分层估计登记原始凭证的差错率为：

$$p_{pst} = \sum_{h=1}^L W_h p_h = 0.7 \times \frac{1}{43} + 0.3 \times \frac{2}{57} = 2.68\%$$

方差为：

$$V(p_{pst}) = \frac{1 - f}{n} \sum_{h=1}^L W_h \frac{n_h}{n_h - 1} p_h(1 - p_h) + \frac{1}{n^2} \sum_{h=1}^L (1 - W_h) \frac{n_h}{n_h - 1} p_h(1 - p_h) = 2.69284 \times 10^{-4}$$

标准差为： $\sqrt{V(p_{pst})} \approx 0.0164$