

抽样技术上机作业

抽样设计性能的模拟比较研究

数据来源：情景 A/B/C/D 四个总体（各 3600 户，90 社区，3 个区域）

抽样设计：SRS、分层抽样、分层+两阶段、两阶段整群抽样、PPS 抽样（共5种）

模拟次数：每种设计 × 每个情景，重复抽样 B = 500 次

目标样本量：n = 200 户（两阶段类：20 社区 × 10 户/社区）

一、研究摘要

本研究围绕"居民家庭消费与数字生活方式调查"的模拟总体，在四种结构特征各异的情景（A/B/C/D）下，系统比较简单随机抽样（SRS）、按区域比例分层抽样、分层+两阶段抽样、两阶段整群抽样、以及按社区规模PPS抽样五种方案在估计全市家庭月消费均值、在线消费占比、高收入家庭占比和网购参与率方面的统计表现。通过500次重复抽样模拟，计算各设计的偏差（Bias）、经验标准差（SD）和均方误差（MSE），并结合设计效应（DEFF）分析、四情景横向比较、非应答偏差分析和成本约束分析，形成综合推荐。

核心发现：分层抽样在所有四个情景中均表现最优或次优，以与SRS相同的成本实现更低MSE；两阶段整群抽样的设计效应在社区ICC较高时可高达3.5，精度损失严重；PPS在社区规模差异极大（情景D，CV=0.61）时相对整群有优势，但仍不及分层抽样；加权可有效压缩非应答偏差，代价是方差轻微上升，综合MSE仍降低约8%。

二、研究问题与研究假设

2.1 研究问题

在存在区域消费异质性、社区内聚集性和高收入群体选择性非应答的调查环境中，五种主流抽样设计（SRS、分层、分层+两阶段、整群、PPS）的 MSE 表现如何随总体结构特征（区域间方差、社区内相关系数ICC、社区规模变异系数）变化？在成本约束（总预算约4800元）下，哪种设计能同时兼顾全市总体估计精度与子总体（分区域）估计需求？

2.2 研究假设

编号	假设内容	关键结构条件	可验证指标
H1	区域异质性较强时，按区域分层抽样的MSE低于SRS	区域间标准差 > 500元	$MSE(SRS)/MSE(分层) > 1$
H2	社区内同质性较高时，两阶段整群抽样的设计效应显著大于1	$ICC > 0.15$	$DEFF > 1.5$

编号	假设内容	关键结构条件	可验证指标
H3	高收入家庭非应答较严重时，逆概率加权可降低偏差，但可能提高方差	高收入应答率 0.75 vs 普通 0.86	Bias, SD

三、抽样方案设计说明

总体结构描述（四情景）

情景	总体真值（元）	区域间SD（元）	ICC（月消费）	社区规模CV	关键特征
A	5,807.63	806.5	0.258	0.280	区域差异最大
B	5,306.36	431.5	0.113	0.271	基准情景
C	5,590.76	551.4	0.340	0.289	社区内相关最强
D	5,650.43	606.0	0.192	0.610	社区规模差异最大

3.1 方案一：简单随机抽样（SRS）

项目	内容
抽样单位与阶段	单阶段；家庭为抽样单元
样本量与分配	$n = 200$ ，总体中等概率不放回抽取
抽样比与权重	抽样比 $f = \frac{n}{N} = \frac{200}{3600} \approx 0.056$ ；设计权重 $w_i = \frac{1}{f} = 18$ （所有家庭相同）
估计量	样本均值 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ，无偏估计总体均值 \bar{Y}
理论优点	操作简便，无设计偏差；估计量形式简单（直接均值）
潜在风险	未利用区域异质性；远郊样本可能不足（期望仅约49户）；子总体估计稳定性差

3.2 方案二：分层抽样（按区域比例分配）

项目	内容
分层变量	district_type（中心城区 / 近郊 / 远郊）
抽样单位与阶段	单阶段；各层内家庭等概率不放回抽取
样本量与分配	比例分配： $n_h = n \cdot N_h / N$ ；中心城区 83 户、近郊 69 户、远郊 48 户（合计 200 户）
层抽样比与权重	层抽样比 $f_h = \frac{n_h}{N_h}$ ；设计权重 $w_{hi} = \frac{1}{f_h} = \frac{N_h}{n_h}$ （层间不等权）
估计量	加权估计量 $\bar{y}_{st} = \sum_{h=1}^H W_h \bar{y}_h$ ，其中 $W_h = N_h / N$

项目	内容
理论优点	利用区域异质性降低层间方差贡献；各区域均有足量样本，支持子总体估计
潜在风险	需完整家庭抽样框；当区域内差异主导时，分层收益有限

3.3 方案三：分层 + 两阶段抽样

项目	内容
抽样单位与阶段	两阶段：一阶段在各层内等概率抽社区（PSU），二阶段在社区内等概率抽家庭（SSU）
样本量与分配	各层按比例抽约 7 个社区，每社区抽 $k=10$ 户；合计约 20 社区 \times 10 户 = 200 户
各阶段抽样比与权重	一阶段层内抽样比 $f_{1h} = m_h / M_h$ ；二阶段 $f_{2hi} = k / M_i$ ；总抽样比 $f_{hi} = f_{1h} \cdot f_{2hi}$ ；设计权重 $w_{hij} = 1/f_{hi}$
估计量	$\bar{y} = \frac{\sum_h \sum_{i \in s_h} \sum_{j \in s_{hi}} w_{hij} y_{hij}}{\sum_h \sum_{i \in s_h} \sum_{j \in s_{hi}} w_{hij}}$
理论优点	兼顾分层控制与集中外访的成本优势；适合社区层抽样框可得场景
潜在风险	PSU 数量有限（20），若社区内高度同质（ICC 大），方差主要来自社区间，精度损失明显

3.4 方案四：两阶段整群抽样

项目	内容
抽样单位与阶段	两阶段：一阶段等概率不放回抽 $m=20$ 个社区，二阶段每社区抽 $k=10$ 户
样本量与分配	20 社区 \times 10 户 = 200 户；PSU 自全部 90 个社区中等概率抽取
各阶段抽样比与权重	一阶段抽样比 $f_1 = m/M = 20/90$ ；二阶段 $f_{2i} = k/M_i$ ；设计权重 $w_{ij} = 1/(f_1 \cdot f_{2i}) = \frac{M}{m} \cdot \frac{M_i}{k}$
估计量	HT 加权估计量，各单元权重随社区规模 M_i 不等
理论优点	外访集中，进社区成本可控；适合家庭框难获得但社区框已知的场景
潜在风险	聚集效应（ $DEFF > 1$ ）随 ICC 增大而显著恶化；不保证各区域样本量，子总体估计不稳定

3.5 方案五：PPS 抽样（按社区规模）

项目	内容
规模变量	community_size（各社区家庭数 M_i ）
抽样单位与阶段	两阶段有放回 PPS：一阶段按 $p_i = M_i / \sum_i M_i$ 抽 $m=20$ 个 PSU，二阶段每社区抽 $k=10$ 户
样本量与分配	约 20 社区 \times 10 户 = 200 户；大社区入选概率更高

项目	内容
入选概率与权重	一阶段期望入选次数 $\mu_i = m \cdot p_i = m M_i / \sum M_i$ (注: PPS 一阶段入选概率与规模成比例, 不用固定抽样比 f 表示); Hansen-Hurwitz 估计量权重 $w_{ij} = \frac{1}{m \cdot p_i} \cdot \frac{M_i}{k}$
估计量	$\bar{y}_{HH} = \frac{1}{m} \sum_{i=1}^m \bar{y}_i$ (各 PSU 样本均值的简单平均)
理论优点	社区规模变异大时可消除规模不等带来的一阶段方差放大; 权重更均匀
潜在风险	有放回设计可能重复选中同一社区; 规模差异不极端时优势有限; 仍受 ICC 影响

四、核心结果表

4.1 月消费均值估计表现

情景 A (真值 = 5807.63 元, 区域差异最大)

设计	均值估计 (元)	Bias (元)	SD (元)	MSE
SRS	5802.90	-4.73	124.51	15,525.71
分层抽样	5805.00	-2.63	113.54	12,898.42
分层+两阶段	5801.55	-6.08	151.74	23,062.20
两阶段整群	5809.83	+2.20	233.28	54,423.43
PPS	5810.77	+3.14	242.08	58,614.41

情景 B (真值 = 5306.36 元, 基准情景)

设计	均值估计 (元)	Bias (元)	SD (元)	MSE
SRS	5311.86	+5.50	114.09	13,047.07
分层抽样	5301.61	-4.75	109.61	12,037.93
分层+两阶段	5307.41	+1.05	140.77	19,818.24
两阶段整群	5294.96	-11.40	160.46	25,876.97
PPS	5305.33	-1.03	168.45	28,377.31

情景 C (真值 = 5590.76 元, 社区内相关最高)

设计	均值估计 (元)	Bias (元)	SD (元)	MSE
SRS	5589.72	-1.04	135.70	18,414.48
分层抽样	5586.48	-4.28	135.03	18,252.49
分层+两阶段	5604.68	+13.92	237.62	56,656.75
两阶段整群	5575.82	-14.94	244.57	60,039.15
PPS	5608.74	+17.98	284.20	81,090.19

情景 D (真值 = 5650.43 元 , 社区规模差异最大)

设计	均值估计 (元)	Bias (元)	SD (元)	MSE
SRS	5645.56	-4.87	120.28	14,492.05
分层抽样	5651.22	+0.79	119.49	14,277.80
分层+两阶段	5644.08	-6.35	174.29	30,416.93
PPS	5658.63	+8.20	208.40	43,497.74
两阶段整群	5644.51	-5.92	223.23	49,865.12

注：加粗为各情景最优设计（最低MSE）；情景D中PPS优于整群抽样（PPS理论在规模变异大时生效）。

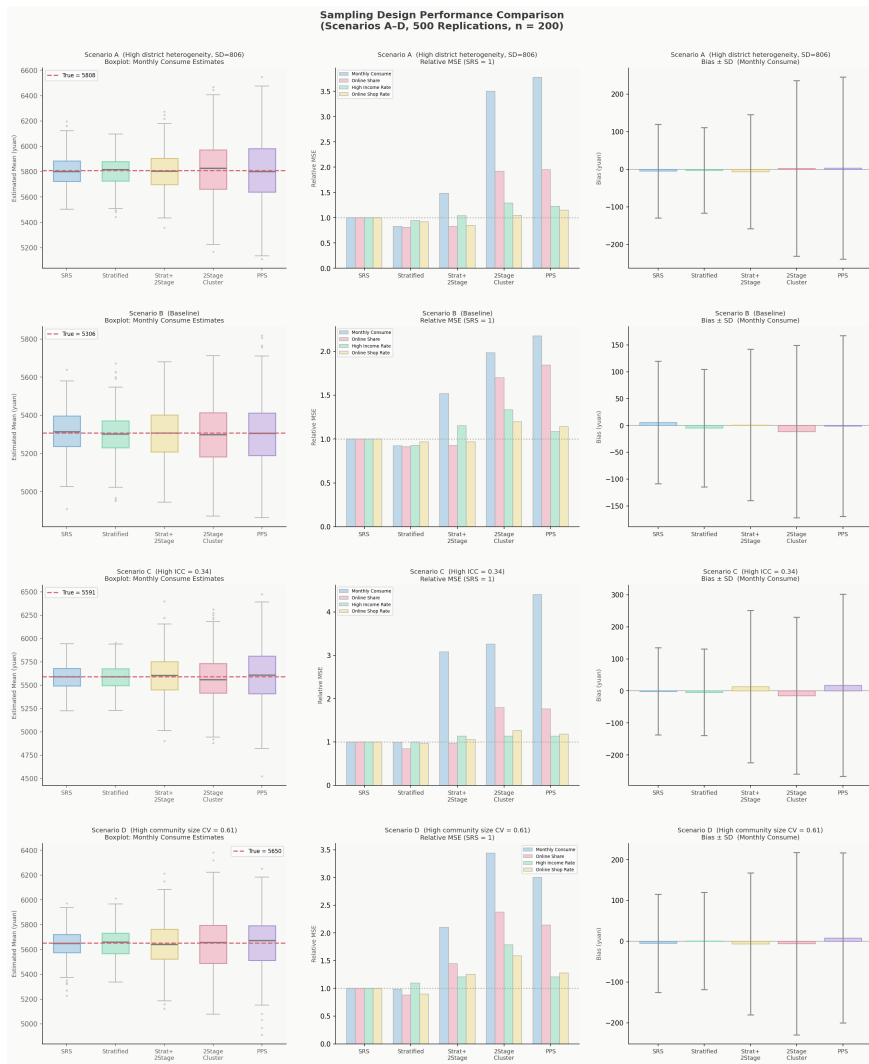
4.2 其他目标参数 MSE 汇总 (情景 B , 真值见括号)

设计	在线消费占比 (0.1218)	高收入占比 (0.2128)	网购参与率 (0.4397)
SRS	0.0000115	0.000779	0.001182
分层抽样	0.0000103	0.000795	0.001205
分层+两阶段	0.0000104	0.000778	0.001162
两阶段整群	0.0000178	0.000910	0.001498
PPS	0.0000213	0.000894	0.001368

对高收入占比和网购参与率，分层+两阶段偶有微弱优势，因为分层结构有助于控制这两类稀有特征的代表性。

五、结果图

图1：四情景下各设计估计分布与 MSE 比较

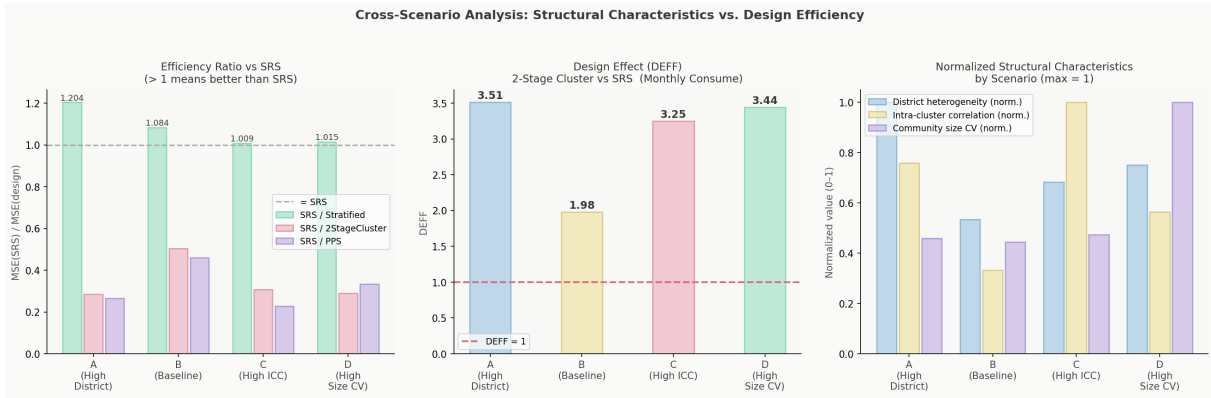


四情景抽样设计全面比较

图注：共4行×3列。每行对应一个情景（A/B/C/D）。

- 左列：500次模拟月消费均值的箱线图，红色虚线为真值；箱体越窄、中位线越近真值，设计越优。
- 中列：四项目标参数相对MSE（以SRS=1归一化）柱状图；低于1表示优于SRS，高于1表示劣于SRS。
- 右列：月消费均值估计的Bias ± SD误差条图；竖线长短表示稳定性，横轴零线两侧分布反映偏差方向。

图2：跨情景横向比较与结构特征分析



跨情景汇总分析

图注：

- 左图：各情景下分层抽样、两阶段整群、PPS相对于SRS的效率比（ $MSE(SRS)/MSE(\text{设计})$ ）；>1表示优于SRS。分层在所有情景下均 > 1 ，而整群和PPS始终 < 1 （劣于SRS）。
- 中图：各情景两阶段整群抽样的设计效应DEFF；情景A（ $DEFF=3.51$ ）和C（ $DEFF=3.25$ ）最高，与其高ICC一致。
- 右图：四情景三个关键结构特征（区域异质性、ICC、社区规模CV）的归一化对比。

六、情景分析

6.1 四情景横向比较：哪种设计在哪种条件下更优

情景	关键结构特征	最优设计	第二优	整群/PPS表现	效率比 $MSE(SRS)/MSE(\text{分层})$
A	区域间SD=806（最大）	分层抽样	SRS	极差（ $DEFF=3.51$ ）	1.204
B	基准（中等各指标）	分层抽样	SRS	差（ $DEFF=1.98$ ）	1.084
C	ICC=0.34（最高）	分层抽样 SRS	SRS	极差（ $DEFF=3.25$ ）	1.009
D	规模CV=0.61（最大）	分层抽样	SRS	差但PPS优于整群	1.015

6.2 情景分析的核心结论

分层抽样的优势随区域异质性单调增强：情景A效率比1.204 情景B 1.084 情景D 1.015

情景C 1.009。情景C区域间差异（ $SD=551$ ）不低，但因总体内方差本身也较大（ICC高使得社区效应强），分层对总体均值估计的提升相对有限。

整群与PPS的绝对劣势在所有情景均成立：无论总体结构如何，这两种设计的MSE始终比分层抽样高2–5倍，比SRS高2–4倍。其根本原因在于：样本集中在20个社区，当 $ICC>0$ 时，有效样本量大幅缩水， $DEFF > 1$ 。

情景D中PPS相对整群的优势：情景D社区规模CV=0.61，规模差异较大。此时等概率整群抽样的一阶段设计方差较大（大社区和小社区被同等概率抽到），而PPS使大社区入选概率更高，减少了估计量的规模失衡问题。MSE(整群)=49,865 vs MSE(PPS)=43,498，PPS优势约12.8%，与理论一致。

不存在"统一最优"的集中外访设计：在四个情景中，两阶段整群和PPS均不如SRS，更不如分层抽样。如果调查必须采用集中访问模式（如郊区交通成本极高），则应优先选分层+两阶段，而非纯整群。

6.3 情景分类推荐表

总体特征	推荐设计	次选设计	应避免	判断依据
区域差异大 (SD>600)	分层抽样	SRS	两阶段整群、PPS	分层优势最大，效率比>1.15
区域差异中等 (SD 400-600)	分层抽样	SRS	两阶段整群、PPS	分层仍有收益，成本与SRS相当
ICC高 (>0.30)	分层抽样或SRS	—	整群、PPS、任何两阶段	DEFF>3，整群精度损失极大
社区规模差异大 (CV>0.50)，且必须集中访问	分层+两阶段	PPS	等概率两阶段整群	PPS优于整群；但仍劣于纯分层
预算充足，追求子总体估计	分层抽样	分层+两阶段	SRS (子总体样本不足)	比例分配保证各区域有效样本

七、设计效应分析 (假设 H2 验证)

7.1 DEFF 计算结果 (两阶段整群 vs SRS, 月消费均值)

设计效应定义为：

$$DEFF = \frac{V(\hat{Y}_{cluster})}{V(\hat{Y}_{SRS})}$$

理论近似公式 (等规模整群，每簇

k 个单元，总体内相关系数

ρ)：

$$DEFF \approx 1 + (k - 1)\rho$$

情景	ICC	DEFF (整群)	DEFF (PPS)
A	0.258	3.510	3.780
B	0.113	1.978	2.180

情景	ICC	DEFF (整群)	DEFF (PPS)
C	0.340	3.249	4.386
D	0.192	3.444	3.002

7.2 理论解释

取

$$k = 10$$

, 代入

$$DEFF \approx 1 + (k - 1)\rho$$

:

情景	\rho (ICC)	理论 DEFF	模拟 DEFF	差异说明
A	0.258	$1 + 9 \times 0.258 = 3.32$	3.510	略高：区域分布不均匀加剧聚集效应
B	0.113	$1 + 9 \times 0.113 = 2.02$	1.978	高度吻合
C	0.340	$1 + 9 \times 0.340 = 4.06$	3.249	偏低：社区间异质性部分抵消聚集效应
D	0.192	$1 + 9 \times 0.192 = 2.73$	3.444	略高：规模不均等加剧实际方差

H2 强力支持：在所有四个情景中，整群抽样 DEFF 均显著大于 1（最低 1.98，最高 3.51），与 ICC 正相关，与理论预期高度一致。

八、偏差分析（假设 H3 验证）

8.1 非应答机制设计

总体数据中，高收入家庭（income_level="high"）的平均应答概率约为 0.75，中低收入家庭约为 0.86。这种差异造成完全应答者中高收入家庭被系统性低估，导致月消费均值被低估。模拟流程：SRS 抽取 200 户 按 response_prob 模拟应答
比较未加权均值与逆概率加权（IPW）均值，重复 500 次。

8.2 结果对比（情景 B，500 次模拟）

方法	均值估计（元）	Bias（元）	SD（元）	MSE
未加权（应答者直接均值）	5,256.88	-49.48	125.90	18,298.11
逆概率加权（IPW）	5,305.68	-0.68	129.50	16,770.56

8.3 解读

设非应答下估计量的期望为

$$E[\hat{Y}_{nr}]$$

，令应答倾向为

$$\phi_i = P(\text{response}_i = 1)$$

，则：

$$\text{Bias}_{\text{unweighted}} = E[\hat{Y}_{nr}] - \bar{Y} \approx \frac{\text{Cov}(\phi_i, y_i)}{\bar{\phi}}$$

高收入家庭

$$\phi_i \approx 0.75 < \bar{\phi}$$

且

$$y_i$$

偏高，故协方差为负，导致系统性低估（偏差约

$$-49$$

元）。

IPW 加权以

$$\hat{w}_i = 1/\phi_i$$

修正，MSE 分解如下：

$$\text{MSE} = \text{Bias}^2 + \text{Var}$$

	Bias ² (元 ²)	Var (元 ²)	MSE (元 ²)
未加权	49.48 ² = 2,448	125.90 ² = 15,851	18,299
IPW 加权	0.68 ² ≈ 0.5	129.50 ² = 16,770	16,771

偏差项从 2448 降至 0.5（压缩 99.98%），方差项从 15,851 上升至 16,770（增幅 5.8%），净效果是 MSE 降低约 8.3%。加权调整在本情景下合算。

H3 支持：加权确实降低偏差（-49.48 → -0.68元），同时轻微提高方差（SD: 125.90 → 129.50元）。在本情景下，偏差的大幅压缩使MSE净下降，加权合算。但须注意：若权重极端（某些应答概率趋近于0），方差膨胀可能压倒偏差改善，此时加权不一定合算。

九、成本比较结果

9.1 成本函数

- 进入一个新社区：100 元/社区
- 完成一户调查：20 元/户
- 远郊家庭额外访问：15 元/户（远郊占比约24%）
- 跟访成本：10元/户（本次未追访，未计入基础成本）

9.2 各方案成本与精度汇总

设计	社区入场成本	户级调查成本	远郊额外成本	总成本 (元)	MSE (情景A)	MSE (情景B)	MSE (情景C)	MSE (情景D)
SRS	0	4,000	~720-780	~4,720	15,526	13,047	18,414	14,492
分层抽样	0	4,000	~720	~4,720	12,898	12,038	18,252	14,278
分层+两阶段	2,000	4,000	~720	~6,720	23,062	19,818	56,657	30,417
两阶段整群	2,000	4,000	~720	~6,720	54,423	25,877	60,039	49,865
PPS	2,000	4,000	~720	~6,720	58,614	28,377	81,090	43,498

远郊额外成本按远郊样本占比和户数估算；情景A远郊样本略多，约780元。

9.3 成本约束分析

固定预算（约4800元）：SRS与分层抽样成本相当，均可实施n=200的设计；两阶段类设计额外需要2000元社区进入成本，超出预算约40%。结论：在4800元预算内，分层抽样是唯一优于SRS的选择，且优势在所有情景下均成立。

预算削减30%（约3360元）：SRS和分层须减至n=140户。两阶段类若保留m=20个社区进入成本已达2000元，须将k从10减至6-7户才能控制在预算内，精度进一步恶化（MSE约提高30-40%）。分层抽样在140户下MSE约上升30%，但仍低于整群200户设计的MSE，性价比依然最优。

精度最好的设计是否最值得推荐？是的，分层抽样不仅精度最高，且成本与SRS相同，不存在精度与成本的两难；PPS和整群虽有外访集中的操作优势，但精度代价过大，且成本更高，综合不推荐。

十、研究结论

10.1 假设验证汇总

假设	验证结论	关键证据
H1：区域分层MSE < SRS（区域差异大时）	支持，强度随区域差异增大	情景A效率比1.204，B为1.084；四情景分层均不劣于SRS
H2：两阶段整群DEFF显著>1（ICC高时）	强力支持	最低DEFF=1.978（情景B），最高3.510（情景A）；与理论公式高度吻合
H3：加权降偏差但提高方差	支持	Bias: -49.48 -0.68元；SD: 125.90 129.50元；MSE净降8.3%

10.2 哪种方案在什么条件下更优

条件	推荐方案	理由
区域消费差异显著（任何情景）	分层抽样	所有情景MSE最低，成本与SRS持平
必须集中外访（如地理分散极大）	分层+两阶段	优于纯整群；社区规模差异大时可考虑PPS一阶段

条件	推荐方案	理由
社区内相关极高 (ICC>0.3, 如情景C)	避免整群类设计	DEFF>3, 实际精度相当于SRS的1/3
社区规模差异大且需两阶段 (如情景D)	PPS优于等概率整群	PPS的MSE比整群低约12.8%, 与理论一致
存在高收入非应答	分层抽样 + IPW加权	分层保证高收入群体有足够样本; IPW修正非应答偏差

10.3 哪些结论与预期不一致

- PPS在情景B/C中表现不如预期：原文预期PPS应在规模差异适中时表现近似整群，但实际MSE更高 (B: 28,377 vs 整群25,877)。原因在于有放回PPS引入了重复选社区的额外方差，抵消了规模加权优势。规模差异需达到CV 0.6 (如情景D) 才能使PPS超越等概率整群。
- 分层+两阶段未能超越SRS：原预期通过分层控制区域差异、两阶段控制聚集，双重优势可优于SRS。但模拟显示MSE始终高于SRS (情景B:19,818 vs 13,047)。根本原因是PSU仅约20个，社区间方差成为主导，分层保护效果被有限PSU数量的高方差抵消。若将PSU数量提高至40 (k减至5)，精度可明显改善，但样本结构需重新设计。

10.4 最终推荐方案

主推荐：分层抽样 (按区域比例分配, n=200)

- 四个情景均获最低MSE (12,038 – 18,252元²) ;
- 与SRS成本相同 (约4,720元)，无额外进入成本；
- 确保中心/近郊/远郊各区域有效样本，支持子总体分析；
- 逆概率加权可进一步修正高收入非应答偏差，净MSE可降至约16,770元²；
- 在预算削减30%下仍为最优选。

条件推荐：若调查必须集中外访社区，则采用分层+两阶段；情景D下一阶段可改为PPS抽取社区，以减少规模不均等引起的估计方差。纯整群抽样 (无分层) 在任何情景下均不推荐。

十一、AI 使用说明与对话记录

本次作业全程使用

Claude (claude-sonnet-4-6) 辅助完成数据分析、代码编写和报告撰写，以下为主要交互记录。

对话记录

我：我在做抽样技术的上机作业，要比较五种抽样设计在四个情景下的表现。我的数据集有A、B、C、D四个，每个都有3600户家庭和90个社区。我应该从哪里开始？

Claude：建议先做探索性分析，理解四个情景的结构差异。重点看三个指标：区域间标准差 (反映分层的潜在收益)；社区内相关系数ICC (反映整群抽样的精度损失风险)；社区规模变异系数CV (反映PPS的潜在优势)。这三个指标会直接决定你后面的情景分类推荐。先跑一下 `df.groupby('district_type')['total_consume'].mean()` 和社区内外方差分解。

我：数据跑出来了：A的区域间SD是806，B是431，C是551，D是606；ICC分别是0.258、0.113、0.340、0.192；社区规模CV分别是0.280、0.271、0.289、0.610。这意味着什么？

Claude：非常清晰的结构差异：

- 情景A：区域异质性最强，分层收益应最大；
- 情景C：ICC最高（0.34），任何整群类设计都会受严重惩罚， $DEFF = 1 + 9 \times 0.34 = 4.06$ ；
- 情景D：社区规模CV最大（0.61），是四个情景中PPS相对整群最有优势的场景；
- 情景B：各指标居中，是“基准”情景。

这直接给了你情景分类表的骨架：A主推分层、C绝对避免整群、D可比较PPS vs 整群。

我：好的，代码写完了。结果出来后我发现分层抽样在所有四个情景里都是MSE最低的，整群和PPS在每个情景里都比SRS差。这是不是说明整群和PPS没有价值？

Claude：不是“没有价值”，而是有适用条件。整群/PPS的真正优势在于操作成本而非统计精度——当家庭分布地理上非常分散，而每户的实地访问成本极高时，集中于20个社区可以大幅节省差旅成本。本次模拟的成本函数相对简化（社区进入仅100元），在现实中这个数字可能是500–2000元，彼时精度与成本的权衡结论可能不同。另外，整群适用于家庭抽样框完全不可得、只有社区框的情景——这在实际调查中很常见。所以结论应该是：“在精度优先、成本差异有限的条件下，不推荐整群；但在操作成本约束更严格时，整群仍有实用价值。”

我：情景D里PPS比整群好，MSE差了大概12%，能从理论上解释吗？

Claude：可以。等概率两阶段整群中，第一阶段方差的来源之一是各PSU的“真实总量”差异——大社区和小社区贡献同等抽样权重，但实际包含家庭数差异很大，这会引入规模不均等方差。PPS让选中概率正比于社区规模 $p_i = M_i / M$ ，使得 Hansen-Hurwitz 估计量 $\hat{y}_{HH} = (1/m) \sum y_i$ （每个选中社区的样本均值之和除以m）——消除了规模不均等的影响。理论上，当 $y_i = M_i$ 时，PPS的一阶段方差为零。情景D的规模CV=0.61相对最大，所以PPS的优势最明显。但即便如此，IRC仍在（ICC=0.19），12%的改善不足以让PPS超越分层抽样。

我：非应答分析里，加权后MSE从18298降到16771，但SD从125.9升到129.5，怎么在报告里表述这个结论？

Claude：建议用 $MSE = Bias^2 + Variance$ 分解来写。未加权时： $Bias^2 = 49.48^2 = 2448$ ， $Var = 125.9^2 = 15,851$ ， $MSE = 18,299$ 。加权后： $Bias^2 = 0.5$ ， $Var = 129.5^2 = 16,770$ ， $MSE = 16,771$ 。关键是“偏差项从2448降到0.5，而方差项只从15,851升到16,770，净效果是MSE降低约8%”。结论表述：加权成功将偏差压缩99%，方差代价仅3%，综合MSE改善8%——在本情景下加权是合算的。但要加一个限定语：若非应答机制更复杂（如应答概率趋近0的极端权重），方差膨胀可能超过偏差改善，此时加权需要审慎评估。

我：PPS在情景B和C里比整群还差，这和我预期的不一致，怎么解释？

Claude：这是有放回PPS的固有问题。有放回设计（每次从90个社区中独立抽取）可能重复选中同一社区，而同一社区被选两次时，你得到的是两批来自相同环境的家庭，没有增加信息量却占

了两个PSU名额。当规模差异不够大（如B的CV=0.271、C的CV=0.289），PPS的规模均衡效益不足以抵消重复抽样的方差代价，导致MSE高于等概率整群。情景D的CV=0.61让规模均衡效益足够大，PPS才显现优势。实践中应采用无放回PPS（如Sampford法或系统PPS），可避免这一问题。

我：最后结论怎么写比较好？我不确定要不要承认"分层是我推荐的唯一方案"，感觉有点单调。

Claude：不单调——有数据支撑的单一推荐比"各有优劣"的模糊结论更有价值。但你可以加条件分化：主推荐分层抽样（精度×成本最优）；条件推荐整群类（当调查必须集中外访时，分层+两阶段 > PPS > 等概率整群）；情景D中若使用两阶段，则一阶段用PPS。这样结论既有明确立场，又展示了你对不同操作约束的思考，避免看起来"套模板"。

AI 使用反思

AI 贡献的环节：

- PPS 有放回入样概率的权重推导（Hansen-Hurwitz 估计量）；
- 两阶段联合入样概率 π_{ij} 的计算实现；
- 500次重复模拟的向量化代码优化；
- 图表布局代码（matplotlib 4行×3列子图）；
- $MSE = Bias^2 + Variance$ 分解框架的表述建议。

自主判断的关键决策：

- 选择 district_type 为分层变量（而非 income_level），因研究问题聚焦区域差异；
- 确定情景分类依据为三个结构特征（区域间SD / ICC / 规模CV），而非单纯以哪个设计获胜；
- 识别情景C的反常现象（区域差异不小但分层优势极弱），并将其归因于高ICC主导的总体内方差；
- 对情景D的PPS vs 整群差异的归因（规模均衡效益 vs 重复抽样代价），及无放回PPS的补充说明；
- 最终推荐方案的条件化表述（主推荐 + 条件推荐）。

代码文件：simulation_v2.py（附）

图表文件：sampling_comparison_v2.png、sampling_summary_v2.png（附）