

附录 1 目标变量 y_{ik} 的生成方式：

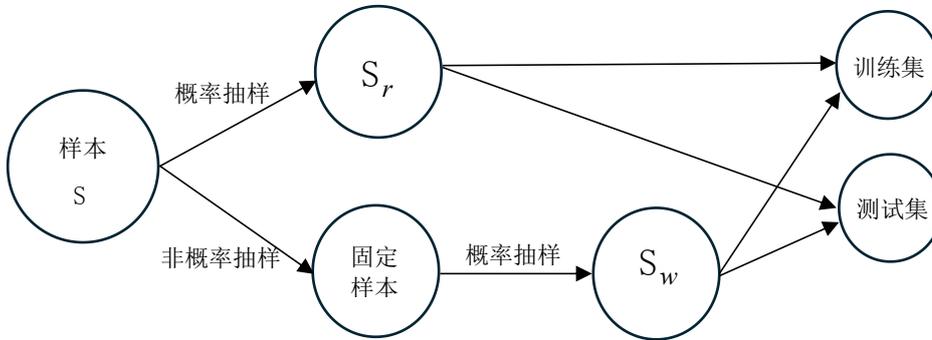
目标变量 y_{ik} 按照以下方式，基于部分协变量 x_{i,k_1} 以及未观测到的影响因素 x_{i,k_1} 生成：

$$y_{i1} = \sum_{k_1} (1 + 0.1k_1)x_{i,k_1} + 1.5q_1 + 1.7q_2 + 1.9q_3 + 2.1q_4 + 2.3q_5 \quad (k_1 = 1, 2, \dots, 10)$$

$$y_{i2} = \sum_{k_2} \sum_{k_1} (1 + 0.1(k_1 + 10k_2))x_{i,k_1+10k_2} + 1.5q_1 + 1.7q_2 + 1.9q_3 + 2.1q_4 + 2.3q_5 \quad (k_2 = 1, 2, \dots, 10; k_1 = 1, 5, 9)$$

$$y_{i3} = \sum_{k_2} \sum_{k_1} (1 + 0.1(k_1 + 10k_2))x_{i,k_1+10k_2} + 1.5q_1 + 1.7q_2 + 1.9q_3 + 2.1q_4 + 2.3q_5 \quad (k_1 = 1, 2, \dots, 10; k_2 = 1, 5, 9)$$

其中，协变量服从正态分布 $x_{i,k_1} \sim N(0, 1 + 0.001k_1)$ ，其他未观测因素 q_i 也服从特定的分布 $q_1 \sim P(1)$ ， $q_2 \sim Exp(1)$ ， $q_3 \sim Exp(1)$ ， $q_4 \sim U(0, 1)$ ， $q_5 \sim \chi^2(1)$ 。



附图 1 样本融合抽样过程图

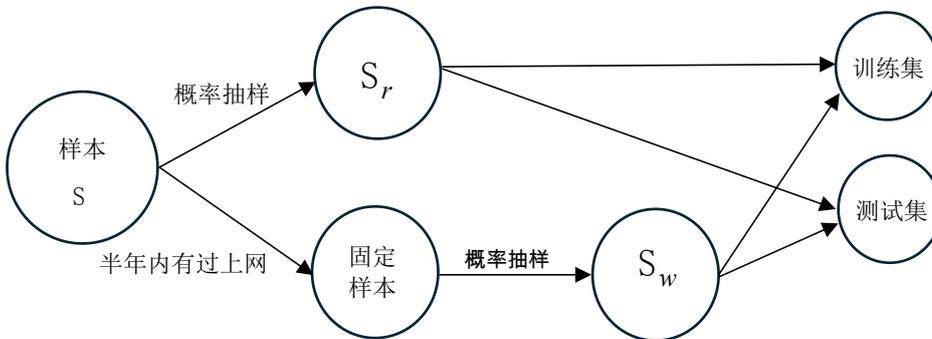
附录 2 入样概率 p_{ik} 的生成方式：

使用回归方程分别在协变量为 30、150、450 个时，为每个样本构造出入样概率 p_{ik} ，并将构造的入样概率作为真实的倾向得分及后续抽样的依据。

$$p_{i1} = \sum_{k_1} (0.5 + 0.1k_1)x_{i,k_1} + 0.5q_1 + 0.3q_3 + 0.1q_5 \quad (k_1 = 2, 4, 6, 8, 10)$$

$$p_{i2} = \sum_{k_2} \sum_{k_1} (0.1(2k_1 + 10k_2))x_{i,2k_1+10k_2} + 0.5q_1 + 0.3q_3 + 0.1q_5 \quad (k_1 = 1, 2, 3; k_2 = 1, 9, 14)$$

$$p_{i3} = \sum_{k_2} \sum_{k_1} (0.1(2k_1 + 10k_2))x_{i,2k_1+10k_2} + 0.5q_1 + 0.3q_3 + 0.1q_5 \quad (k_1 = 1, 2, 3; k_2 = 1, 10, 20, 30, 40)$$



附图 2 实际操作过程中的抽样过程图