

1. 从 CGSS2010 中，选择 4 个变量， y 应为二元变量， x_1 为连续型变量， x_2 为二元变量， x_3 为类别变量，进行 logistic 回归分析。
 - 请说明你关心的研究问题为何？并以一句话或一个理论说明你的预期为何？
 - 这些变量中，何为自变量？何为因变量？
 - 对变量进行缺失值处理（请剔除，不要插补）和适当的重新编码。
 - 进行 logistic 回归分析 $\Pr(y = 1) = \text{logit}^{-1}(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_{3:1} + \beta_4 x_{3:2})$ （假设 x_3 三个类别的定类变量，系数为 β_3, β_4 ，如果 x_3 类别大于三类，则将有更多系数），汇报回归截距 α 、回归系数 $\beta_1, \beta_2, \beta_3, \beta_4$ 为何？
 - 进行偏差和似然比检验的 χ^2 检验。
 - 回归 Psuedo R^2 为何？如何解读 Psuedo R^2
 - 使用 Z 检验 $H_0: \beta_i = 0$ ，并计算各个 β_i 的 95% 信用区间。
 - 解读回归系数（这里务必每个系数都详细解读）。
 - 使用散点图画出 y 和 x_1 的关系，并依照 x_2 的取值，画出两条回归曲线，将 x_3 固定于其中一类。
 - 考虑至少一组交叉项，并重复以上步骤！