

## Appendix 1

```

setwd("D:/R work")
#### 1.设定发病率:
# 输入估算的发病率
prevalence <- 0.06

#### 2.计算 max(R2_cs):
# 使用公式计算 max(R2_cs)
# lnL_null: 仅包含截距的模型的对数似然值
# 计算 零模型 (null model) 的对数似然值 lnL_null, 即只有截距的模型的对数似然
lnL_null <- prevalence * log(prevalence) + (1 - prevalence) * log(1 - prevalence)
lnL_null
# 计算 Cox-Snell R2 的理论最大值
# 计算 max(R2_cs)
max_R2_cs <- 1 - exp(2 * lnL_null)

# 输出计算出的 max(R2_cs) 值
cat("The calculated max(R2_cs) value is:", max_R2_cs, "\n")

#### 3.计算期望的 R2_cs:
# 如果我们保守假设新模型将解释 15% 的变异性,
# 则根据此假设计算期望的 R2CS 值:
rsquared <- 0.15 * max_R2_cs # 基于假设的期望 R2CS 值

# 输出期望的 R2CS 值
cat("The expected R2CS value (rsquared) is:", rsquared, "\n")

#### 4.使用 pmsampsize 计算样本量
# 加载必要的库
library(pmsampsize)

# 使用 pmsampsize 函数计算所需的样本量
pmsampsize(
  type = "b", # 'b' 代表二元 (binary) 结局变量
  csrsquared = rsquared, # 使用计算得到的期望 R2CS 值
  parameters = 15, # 候选预测变量的数量 (共 15 个)
  prevalence = prevalence, # 研究人群中结局事件的发生率
  seed = 123456 # 设置随机种子以保证结果可复现
)

# 每个预测变量所需的事件数 (EPP)
# Criteria 1 中的样本量是 2391, 该标准考虑了 收缩系数为0.90 和 EPP要求为9.56, 旨在确保模型有足够的样本量来进行有效的拟合和减少过拟合。
# Criteria 2 中的样本量为 792, 该标准假设 0.75的收缩系数 和较低的 EPP (3.17), 使得样本量较小, 适用于降低预测误差。
# Criteria 3 中的样本量为 87, 这是最小样本量要求, 但其 EPP为0.35, 可能导致预测效果不稳定。

```