2 疫苗有效性(Vaccine Efficacy)的样本量计算

我们首先导入后面会用到的包，后续会用到dplyr的%>%管道函数和kableExtra输出表格。

```
library(dplyr)
library(kableExtra)
```

2.1 疫苗有效性(Vaccine Efficacy)

- 当我们检验一款新的疫苗是否有效时，通常会使用疫苗有效性π(Vaccine Efficacy)这个指标。
- 让P_1, P_2分别为安慰剂组和接种疫苗组的发病率；N_1, N_2分别为安慰剂组和接种疫苗组的群体总数；X, Y分别为安慰剂和疫苗组的病例人数。则$X \overset{i.i.d}{\sim} \text{Binomial}(N_1, P_1)$, $Y \overset{i.i.d}{\sim} \text{Binomial}(N_2, P_2)$。
- 疫苗有效性可以可以用可以被写成：

$$\pi = 1 - (P_2/P_1) \tag{2.1}$$

- 原假设和备择假设可以写成

$$H_0 : \pi \leq \pi_0 \text{ versus } H_1 : \pi > \pi_0 \tag{2.2}$$

2.2 基于泊松假设的大样本

- 如果一个疾病发病率较低，我们就需要更多受试者参加实验，在这种情况下X, Y可以被近似成为独立的泊松分布.
- $X \overset{i.i.d}{\sim} \text{Poisson}(\lambda_1)$, $Y \overset{i.i.d}{\sim} \text{Poisson}(\lambda_2)$ with $\lambda_1 = N_1 \cdot P_1, \lambda_2 = N_2 \cdot P_2$.
- Y的条件概率分布可以被写成 $Y|T \sim \text{Binomial}(T, \theta) = \binom{T}{k}\theta^k(1-\theta)^{T-k}$ with $T = X + Y, \theta = \frac{\lambda_1}{\lambda_1 = \lambda_2}$
- 此时原假设和备择假设可以写成

$$H_0 : \theta \geq \theta_0 \text{ versus } H_1 : \theta < \theta_0 \text{ where } \theta_0 = \frac{1-\pi_0}{2-\pi_0} \tag{2.3}$$

- P value 可以被计算为

$$p = \Pr[Y \leq Y_{\text{obs}} \mid Y \sim \text{Binomial}(T, \theta_0)] = \sum_{k=0}^{Y_{\text{obs}}} \binom{T}{k} \theta_0^k (1-\theta_0)^{T-k} \tag{2.4}$$

- Statistical power 可以被计算为

$$1 - \beta = \Pr[Y \leq Y_c \mid Y \sim \text{Binomial}(T, \theta_1)] = \sum_{k=0}^{Y_c} \binom{T}{k} \theta_1^k (1-\theta_1)^{T-k} \tag{2.5}$$

- 临床实验所需样本量可以被计算为

$$N_2 = T/[(2-\pi_1)/P_1] \tag{2.6}$$

- 计算样本量所需的变量为π_0, π_1, α, 期望统计功效(1-β)和安慰剂组发病率(P_1)。

2.3 R代码示例

```r
exact_conditonal_test <- function(T, alpha, power, theta0, theta1) {
  Y_c <- qbinom(alpha, size = T, prob = theta0)-1
  p_value <- pbinom(Y_c, size = T, prob = theta0)
  power <- pbinom(Y_c, size = T, prob = theta1)
  return(c(T, Y_c, power, p_value))
}

ect_sample_size <- function(T, pi0, pi1, incidence, alpha, power){
  theta0 <- (1-pi0) / (2-pi0)
  theta1 <- (1-pi1) / (2-pi1)
  table_t <- as.data.frame(do.call(rbind,
                                   lapply(T, exact_conditonal_test, alpha = alpha, theta0= theta0, theta1 = theta1)), .name_repair =
"unique")
  colnames(table_t) <- c('T', 'Y_c', 'power', 'p-value')

  for (n in 1:nrow(table_t)){
    if (table_t$power[n] >= power && all(table_t$power[n:nrow(table_t)] >= power)) {
      min_n <- table_t$T[n]
      break
    }
  }
  result <- list(
    T_table = table_t,
    T = min_n,
    text = paste0('The min value of T achieve power of ', power, ' is ', min_n)
  )
  class(result) <- 'result'
  result
}

print.myresult <- function(x, ...) {
  cat(x$text, "\n")
}


T2N <- function(T_value, incidence, pi1, dropout_rate){
  N2 <- T_value/((2-pi1)*incidence)/(1-dropout_rate)
  cat('The sample of vaccine group considering the drop out rate:', N2)
}
```

详细推导过程可以看 Ivan S. F. Chan and Norman R. Bohidar[1] 和 Matthew M Loiacono[2]

2.3.1 例1

Chan and Bohidar[3] 这篇论文介绍了exact conditional method的公式推导。他还给了一个不同样本量对应的power和significance level。 我们可以运行上面的函数来验证结果的准确性。设置病例范围从33到40，$\pi_0 = 0.2, \pi_1 = 0.8, P_1 = 0.006, \alpha = 0.025$, 期望统计功效为95%。输入以上参数到ect_sample_size这个函数。

Table II
Sample Size Determination Using the Exact Conditional Test Based on Poisson Assumption

Total Number of Cases (T)	Critical Value (Y_c)	Exact Power (%)	Exact Level (%)
33	8	91.4	1.36
34	9	95.4	2.44
35	9	94.5	1.79
36	9	93.4	1.30
37	10	96.5	2.28
38	10	95.8	1.68
39	10	95.0	1.23
40	11	97.4	2.11

Note: The hypothesis H_0: $\pi \leq 0.2$ versus H_1: $\pi > 0.2$ is tested at the nominal 2.5% level. A true efficacy of 0.8 is assumed under the alternative.

```
T <- 33:40 # the number of T
pi0 <- 0.2 # null hypothesis efficacy
pi1 <- 0.8 # True efficacy under alternative hypothesis
alpha <- 0.025 # type I error
incidence <- 0.006 # placebo incidence rate
power <- 0.95
res1 <- ect_sample_size(T, pi0, pi1, incidence, alpha, power)
kbl(res1$T_table)%>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

T	Y_c	power	p-value
33	8	0.9139690	0.0136117
34	9	0.9540856	0.0244451
35	9	0.9449925	0.0178969
36	9	0.9347919	0.0129998
37	10	0.9653937	0.0227940
38	10	0.9584044	0.0168288
39	10	0.9504998	0.0123313
40	11	0.9738542	0.0211901

代码中的incidence 就是P_1， 其他参数与上面提及的保持一致。可以看到输出的表格中样本量，critical value， statistical power和p value与论文中的表格完全一致。我们打印出能够达到95% statistical power的病例数

```
res1$T
#> [1] 37
```

打印出T=37与论文一致，再用公式(2.6)计算出疫苗组样本量，为了方便计算我也写成了函数t2n。此示例计算样本量时没有考虑脱落所以dropout_rate 参数可以设置为0。得到结果样本量至少为5138.889。这个结果乘2就是疫苗组和安慰剂组所需的总样本量，总样本量至少为10277.78，进一10278。此结果与论文一致，证明我们的算法无误。

```
T2N(37, incidence, pi1, dropout_rate = 0)
#> The sample of vaccine group considering the drop out rate: 5138.889
```

2.3.2 HRV-三期

For the primary hypothesis, HRV was considered efficiency of $>55\%$ (Li et al., 2014; Mo et al., 2017) against any severity of RVGE caused by G1, G2, G3, G4, G8, G9 serotype; the incidence rate of RVGE in two rotavirus seasons was 5%, the ratio of subjects in HRV and placebo group was 1:1; and at least 73 cases of acute gastroenteritis (AGE) of any severity caused by G1, G2, G3, G4, G8, G9 serotype of rotavirus are expected to be observed; HRV had a protective efficacy $>70\%$ for severe RVGE, the cumulative incidence rate of severe RVGE in two consecutive rotavirus seasons was estimated at 1% (Chen et al., 2019; Liu et al., 2020; Zhang et al., 2020); About 20% dropout rate was considered. The sample size was calculated as 6400 subjects using the exact condition method of Chan and Bohidar under the assumption of large sample Poisson distribution (Chan and Bohider, 1998).

Zhiwei Wu et al.[4] 这篇论文原为我们想要复现的样本量。但是过程中发现其中并未明确提及所需参数 π_1, α和期望统计功效。无法计算出样本量所以后续我们使用这篇论文作为参考。下一个示例为HRV-三期这篇论文的参考文献，也是rotavirus vaccine的疫苗有效性临床实验。

2.3.3 RV5

Zhaojun Mo et al.[5] 这篇论文会作为我们后续疫苗有效性检验样本量计算的参考文章。提供参数：15%脱落率，$\pi_0 = 0, \pi_1 = 0.6, P_1 = 0.02, \alpha = 0.025$，期望统计功效为80%。带入以上参数可以得到病例总数至少为47，我们取双数48。再用T2N函数计算疫苗组样本量总数至少2016.807，我们取整为2020，与论文一致。

The study design was RVGE case driven. Under the following assumptions: an 85% evaluability rate, a 60% true efficacy against severe RVGE, and a true underlying attack rate of 2% for severe RVGE, a total of 48 targeted severe RVGE cases, of which no more than 16 are in the RV5 group, would be able to demonstrate the efficacy of RV5 against severe RVGE with 1-sided $\alpha = 0.025$ and power≈80%. To accrue the targeted severe RVGE cases, a total of 4040 participants with 2020 per vaccination group were sufficient. This sample size would be able to demonstrate the efficacy of RV5 against any-severity RVGE with 1-sided $\alpha = 0.025$ and power >90%, from which 100 targeted any-severity RVGE cases would be accrued.

```
T <- 40:50 # the number of T
pi0 <- 0 # null hypothesis efficacy
pi1 <- 0.6 # True efficacy under alternative hypothesis
alpha <- 0.025 # type I error
incidence <- 0.02 # placebo incidence rate
power <- 0.8
res_rv5 <- ect_sample_size(T, pi0, pi1, incidence, alpha, power)
kbl(res_rv5$T_table)%>%
  kable_styling(bootstrap_options = "striped", full_width = F, position = "left")
```

T	Y_c	power	p-value
40	13	0.7692914	0.0192387
41	13	0.7363326	0.0137666
42	14	0.8052771	0.0217793
43	14	0.7757295	0.0157697
44	15	0.8362319	0.0243834
45	15	0.8100042	0.0178489
46	15	0.7819032	0.0129480
47	16	0.8396107	0.0199930
48	16	0.8146130	0.0146525
49	17	0.8650285	0.0221921
50	17	0.8429717	0.0164196

```
res_rv5$T
#> [1] 47
```

```
T2N(48, incidence, pi1, dropout_rate = 0.15)
#> The sample of vaccine group considering the drop out rate: 2016.807
```

Reference

Chan, Ivan S. F., and Norman R. Bohidar. "Exact Power and Sample Size for Vaccine Efficacy Studies." *Communications in Statistics - Theory and Methods* 27, no. 6 (1998): 1305–22. https://doi.org/10.1080/03610929808832160.

Loiacono, Matthew M. *SAMPLE SIZE ESTIMATION AND POWER CALCULATIONS FOR VACCINE EFFICACY TRIALS FOR EXCEEDINGLY RARE DISEASES*. n.d.

Mo, Zhaojun, Yi Mo, Mingqiang Li, et al. "Efficacy and Safety of a Pentavalent Live Human-Bovine Reassortant Rotavirus Vaccine (RV5) in Healthy Chinese Infants: A Randomized, Double-Blind, Placebo-Controlled Trial." *Vaccine* 35, no. 43 (2017): 5897–904. https://doi.org/10.1016/j.vaccine.2017.08.081.

Wu, Zhiwei, Qingliang Li, Yan Liu, et al. "Efficacy, Safety and Immunogenicity of Hexavalent Rotavirus Vaccine in Chinese Infants." *Virologica Sinica* 37, no. 5 (2022): 724–30. https://doi.org/10.1016/j.virs.2022.07.011.

« 1 前言

"**Biostatistical method with R code**" was written by Shuai Zhu. It was last built on 2025-07-23.

This book was built by the bookdown R package.