# 假设检验与两类错误

**Hypothesis testing and two type of errors**

Qingyao Zhang

2026-03-21

## 目录



```
library(tidyverse)
library(pwr)
```

假设检验是进行推断统计的一般思路。假设检验包括两步，第一步是假设，第二步是检验。假设就是对总体的均值和标准差作出假设。检验就是将实际样本放在假设总体的抽样分布中进行检验。下面举例说明。

## 1 抽样

首先对**实际总体**及其样本进行如下设定：

```
# 实际总体的均值
miuH1 <- 103
# 实际总体的标准差
sigma <- 15
```

```r
# 实际样本的样本量
n <- 100
```

某协会有会员 10000 人，其智商 IQ 的均值为 103，标准差为 15。这 10000 人就是一个总体。本文使用 `rnorm()` 函数生成这一**实际总体**的数据。

```r
# 设定随机种子值以保证结果可重复
set.seed(20250530)
# 将 ID 与 IQ 合并在数据框 dat_population 中
dat_population <- data.frame(
  # 10000 人的编号 ID
  ID = seq(1, 10000, 1),
  # 10000 人的 IQ 得分
  IQ = rnorm(10000, miuH1, sigma)
  )
```

从上述总体中随机抽取一个样本量 $n$ 为 100 的样本：

```r
dat_sample <- dat_population |>
  slice_sample(n = n)
```

对该样本进行描述性统计：

```r
# 样本的 IQ 的均值
M_IQ_SAMPLE <- round(mean(dat_sample$IQ), 2)
# 样本的 IQ 的标准差
SD_IQ_SAMPLE <- round(sd(dat_sample$IQ), 2)
paste0(" 该样本 IQ 的均值为", M_IQ_SAMPLE, "，标准差为", SD_IQ_SAMPLE,"。")
```

```
[1] "该样本 IQ 的均值为 103.73，标准差为 14.5。"
```

总体通常是未知的，样本通常是已知的。前文提到的某协会的 10000 名会员那个总体的均值和标准差通常是未知的，而现实中我们可以随机抽取一个 100 人的样本，然后通过这个实际样本来推断其来自的**实际总体**。

## 2 推断统计

前文得到，实际样本的 IQ 的均值为 103.73，标准差为 14.5。我们知道，IQ 作为一个标准分，其均值为 100，标准差为 15。均值 100 的含义是全人类的平均智商为 100。本文提出以下问题：前文抽取的样本的

IQ 的均值与全人类的平均智商 100 是否具有显著差异？该样本的 IQ 的均值在统计上是否等于 100？该样本的 IQ 的均值在统计上是否高于 100？

为回答上述问题，我们需要按照假设检验的思路对样本的 IQ 进行单样本 $z$ 检验或者单样本 $t$ 检验。单样本 $z$ 检验与单样本 $t$ 检验的作用是将**实际样本**的均值与一个**假设总体**的均值进行比较。

## 3 假设：对假设总体与实际总体的均值作出假设

上述问题中，本文实际上关注的是 IQ 的水平的差异，即，样本的均值 103.73 与**假设总体**的均值 100 是否有显著差异，因此**假设总体**的均值为 100。注意，前文提到 100 人的**实际样本**所来自的**实际总体**的均值为 103，标准差为 15。这里本文提出了一个**假设总体**，这个**假设总体**的均值为 100。这个**假设总体**是实际样本用以比较的对象。

**实际样本**的均值是**实际总体**的均值的无偏估计，我们将**实际样本**的均值与**假设总体**的均值进行比较，实际上也就是将**实际样本**所代表的**实际总体**的均值与**假设总体**的均值进行比较。

通常我们并不关注标准差的差异，因此我们假定**实际总体**与**假设总体**的标准差是相同的，都是 15。另外，我们假定**实际总体**与**假设总体**都服从正态分布。

## 4 单样本 $z$ 检验：将实际样本的均值与假设总体的均值进行比较

将实际样本的均值与**假设总体**的均值进行比较，逻辑上是假设实际样本的均值来自**假设总体**，即假设实际样本的均值与**假设总体**的均值在统计上是相同的，即没有显著差异。这一假设被称为虚无假设，又称为零假设，表示为 H0。“零” 即表示实际样本的均值与**假设总体**的均值的差异为零。而我们对实际样本所来自的**实际总体**的假设被称为备择假设，又称为 H1 表示，1 与 0 相对，表示存在差异。

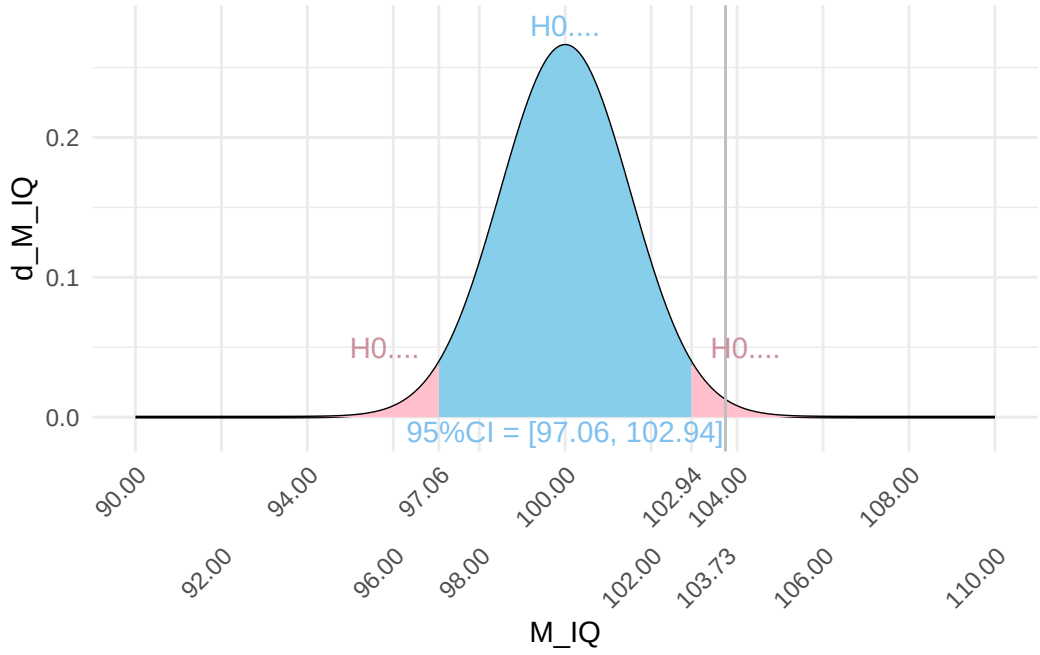
如何检验这一假设呢？我们采用单样本 $z$ 检验将样本均值 103.73 与假设总体的均值 100 进行比较，这本质上是将实际样本的均值放在**假设总体**的均值的抽样分布中，继而判断实际样本的均值在**假设总体**的均值的抽样分布中的位置。统计上以**小概率**为异常，如果 103.73 确实是来自于均值为 100 的总体，那么 103.73 应**大概率**落在 100 的附近，而不应落在远远偏离 100 的位置（极左或极右）。

什么是小概率？多小算小？统计上武断地以 5% 为标准。如果 103.73 是来自于均值为 100 的总体，那么 103.73 应大概率落在 100 的附近，这里的 “大概率” 为 95%。如下图所示，**假设总体**的抽样分布以 100 中心，正态曲线与横坐标轴包围的面积表示累积概率，从横坐标的负无穷大到正无穷大，累积概率为 1。中间的累积概率为 95% 的范围（下图蓝色部分）被称为 100 的 95% 置信区间 (95% Confidence Interval, 95% CI)。如果 103.73 是来自于均值为 100 的总体，那么 103.73 应落在这一 95% 置信区间内。反之，如果 103.73 并非来自于均值为 100 的总体，那么 103.73 应远离抽样分布的中心、落在双侧尾部范围内，左侧尾部的累积概率为 2.5%，右侧尾部的累积概率为 2.5%，两侧尾部的累积概率合计 5%（下图粉色部分）。落在双侧尾部范围内，意味着 103.73 发生的概率小于 5%，这意味着**在 H0 成立时，H1 非常不可能发生**，在下结论时，我们会粗糙地下结论：$p < 0.05$ 意味着**在 H0 成立时，H1 不会发生**。这一结论一定是正确的吗？不一定。“非常不可能发生” 不等于 “不会发生”，在 H0 成立时，H1 仍有 0.05 的概率会发

生，但我们的结论是 H1“不会发生”，这一结论有可能是错的，错误的概率等于 H1 发生的概率，即 0.05。我们将这种错误称为一类错误（Type 1 error），用 $\alpha$ 表示。

```r
# 假设总体抽样分布的作图数据
datH0 <- data.frame(
  # 横坐标为 IQ 均值
  M_IQ_H0 = seq(90, 110, 0.01),
  # 纵坐标为相应的 IQ 均值的概率
  d_M_IQ_H0 = dnorm(seq(90, 110, 0.01), 100, sigma/sqrt(n))
)
# 95%CI 下限
LLCI_H0 <- round(qnorm(0.025, 100, sigma/sqrt(n)), 2)
# 95%CI 上限
ULCI_H0 <- round(qnorm(0.975, 100, sigma/sqrt(n)), 2)
# 95%CI 作图数据
datH095CI <- datH0[(datH0$M_IQ_H0 >= LLCI_H0) & (datH0$M_IQ_H0 <= ULCI_H0), ]
# 左侧尾部作图数据
datH0LeftTail <- datH0[datH0$M_IQ_H0 <= LLCI_H0, ]
# 右侧尾部作图数据
datH0RightTail <- datH0[datH0$M_IQ_H0 >= ULCI_H0, ]
# 作图
H0plot <- ggplot(datH0, aes(M_IQ_H0, d_M_IQ_H0)) +
  geom_path() +
  geom_area(data = datH095CI, fill = "skyblue") +
  geom_area(data = datH0LeftTail, fill = "pink") +
  geom_area(data = datH0RightTail, fill = "pink") +
  geom_vline(xintercept = M_IQ_SAMPLE, color = "gray") +
  geom_text(aes(label = label),
            color = "pink3",
            data = data.frame(
            M_IQ_H0 = c(95.8, 104.2),
            d_M_IQ_H0 = c(0.05, 0.05),
            label = c("H0 左侧尾部", "H0 右侧尾部"))
            ) +
  geom_text(aes(label = label),
            color = "skyblue2",
            data = data.frame(
            M_IQ_H0 = 100,
            d_M_IQ_H0 = -0.01,
            label = paste0("95%CI = [", LLCI_H0, ", ", ULCI_H0, "]"))
            ) +
  geom_text(aes(label = label),
            color = "skyblue2",
```

```r
            data = data.frame(
            M_IQ_H0 = 100,
            d_M_IQ_H0 = 0.28,
            label = "H0 抽样分布")
            ) +
  scale_x_continuous(
    name = "M_IQ",
    breaks = c(90, 92, 94, 96,
               LLCI_H0,
               98, 100, 102,
               ULCI_H0,
               M_IQ_SAMPLE,
               104, 106, 108, 110),
    guide = guide_axis(angle = 45, n.dodge = 2),
    minor_breaks = NULL) +
  scale_y_continuous(name = "d_M_IQ") +
  theme_minimal()
H0plot
```



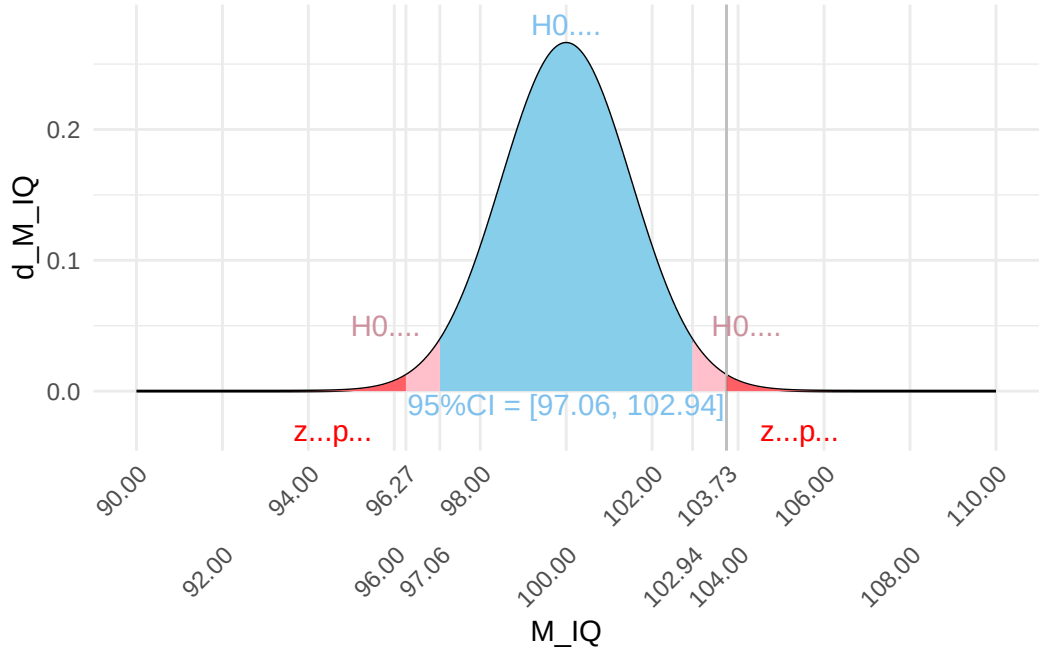

在上图中，实际样本的均值 103.73 落在假设总体抽样分布的右侧尾部。我们可以通过 $z$ 检验计算得到均值 103.73 在假设总体抽样分布中的精确位置及其精确尾部概率 $p$ 值。**假设总体**的抽样分布的中心为 100，标准误为 `sigma/sqrt(n)`。根据公式可得:

```r
z <- (mean(dat_sample$IQ) - 100)/(sigma/sqrt(n))
pz <- (1 - pnorm(z))*2
paste0("*z* =", z, " , *p* =", pz, "。")
```

[1] "$z$ =2.48914419284982 , $p$ =0.0128051019713626。"

103.73 在假设总体抽样分布中的精确位置及其尾部概率见下图:

```r
# z 的精确 p 值左侧尾部作图数据
datH0PzLeftTail <- datH0[datH0$M_IQ_H0 <= (200 - M_IQ_SAMPLE), ]
# z 的精确 p 值右侧尾部作图数据
datH0PzRightTail <- datH0[datH0$M_IQ_H0 >= M_IQ_SAMPLE, ]
H0plot2 <- H0plot +
  geom_area(data = datH0PzLeftTail, fill = "red", alpha = 0.5) +
  geom_area(data = datH0PzRightTail, fill = "red", alpha = 0.5) +
  geom_text(aes(label = label),
            color = "red",
            data = data.frame(
              M_IQ_H0 = c(198.3 - M_IQ_SAMPLE, M_IQ_SAMPLE + 1.7),
              d_M_IQ_H0 = c(-0.03, -0.03),
              label = c("z 的精确 p 值左尾", "z 的精确 p 值右尾"))
            ) +
  scale_x_continuous(
    name = "M_IQ",
    breaks = c(90, 92, 94, 96,
               200 - M_IQ_SAMPLE,
               LLCI_H0,
               98, 100, 102,
               ULCI_H0,
               M_IQ_SAMPLE,
               104, 106, 108, 110),
    guide = guide_axis(angle = 45, n.dodge = 2),
    minor_breaks = NULL
    )
H0plot2
```




可见，103.73 的双尾 $p = 0.0128051$‘。在上图中，103.73 的双尾 $p$ 值为两侧粉色区域累积概率之和。

## 5 单样本 $t$ 检验

总体的标准差通常是未知的，实际总体是这样，假设总体也是这样，因此，一般而言，我们会通过样本的无偏标准差估计**假设总体**的标准差，我们在此基础上得到的统计量不再是 $z$，而是 $t$，这一检验被称为单样本 $t$ 检验。根据 $t$ 的公式可得:

```
t <- (mean(dat_sample$IQ) - 100)/(sd(dat_sample$IQ)/sqrt(n))
pt <- (1 - pt(t, df = n - 1))*2
paste0("*t* = ", t, ", *p* = ", pt, "。")
```

[1] “$t = 2.57529305041762$, $p = 0.0114941607307264$。”

我们也可以使用 `t.test()` 函数进行单样本 $t$ 检验:

```
t.test(dat_sample$IQ, mu = 100)
##
##  One Sample t-test
##
## data:  dat_sample$IQ
## t = 2.5753, df = 99, p-value = 0.01149
```

```
## alternative hypothesis: true mean is not equal to 100
## 95 percent confidence interval:
##  100.8570 106.6105
## sample estimates:
## mean of x
##  103.7337
```

上述单样本 $t$ 检验的结果得到的 $p$ 值与单样本 $z$ 检验是接近的，结论与单样本 $z$ 检验是一致的。

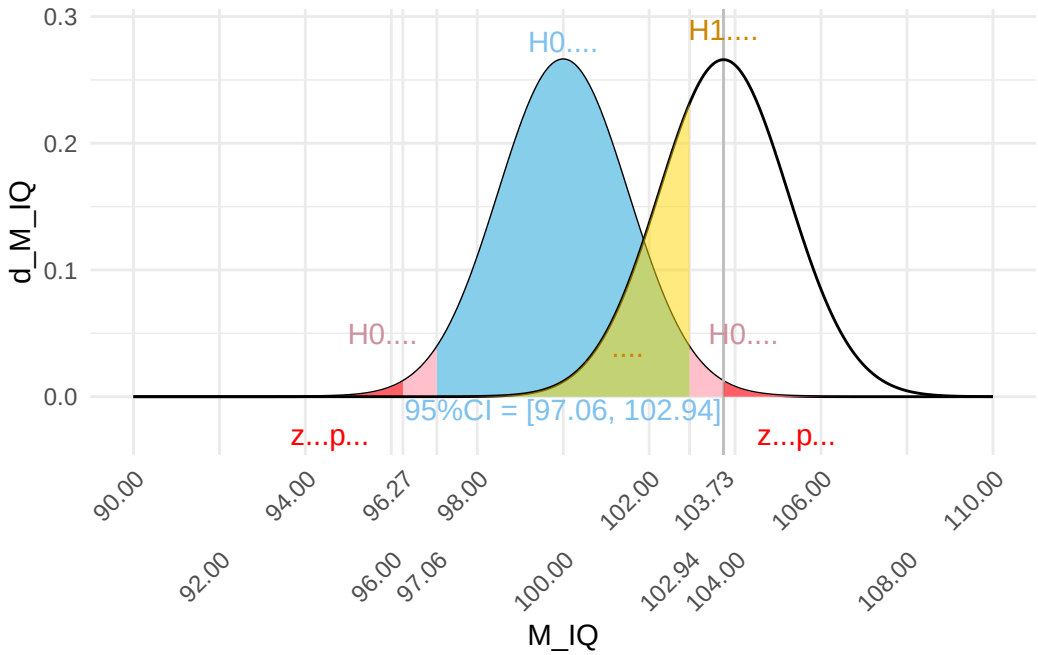
## 6 二类错误

我们在进行推论统计时，总是以**假设总体**的抽样分布为根据，统计分析的目的在于证伪或者否定**假设总体**。这一统计思路与证伪主义的哲学立场一致。所以，前文采用单样本 $z$ 检验或者单样本 $t$ 检验比较实际样本的均值与**假设总体**的均值时，我们都是以**假设总体**为靶子，首先假设**假设总体**为真，然后力图通过统计检验证明**假设总体**为假。我们的统计依据是一类错误 alpha：如果实际样本的均值落于双侧尾部，即，非常不可能发生的事情竟然发生了，那么我们得出结论：我们最初的**假设总体**不成立，H0 不成立，H0 为假。

接下来讨论的“二类错误”的问题稍微复杂一点。前文提到，我们将**实际样本**的均值与**假设总体**的均值进行比较，实际上也就是将**实际样本**所代表的**实际总体**的均值与**假设总体**的均值进行比较。**实际总体**的均值一定是 103.73 吗？不一定。**实际总体**的均值服从抽样分布，**实际总体**的均值落于一定的范围之内，我们将**实际总体**的均值的抽样分布称为 H1 的抽样分布，见下图。

```
# 横坐标为 IQ 均值
M_IQ_H1 <- c(seq(90, 110, 0.01))
# 纵坐标为相应的 IQ 均值的概率
d_M_IQ_H1 <- dnorm(M_IQ_H1, M_IQ_SAMPLE, sigma/sqrt(n))
# 假设总体抽样分布的作图数据
datH1 <- data.frame(M_IQ_H1, d_M_IQ_H1)
# beta
datBeta <- datH1[datH1$M_IQ >= LLCI_H0 & datH1$M_IQ <= ULCI_H0, ]
# 作图
H0H1plot <- H0plot2 +
  geom_path(aes(M_IQ_H1, d_M_IQ_H1), data = datH1) +
  geom_area(aes(M_IQ_H1, d_M_IQ_H1), data = datBeta,
            fill = "gold", alpha = 0.5) +
  geom_text(aes(x = M_IQ_H1, y = d_M_IQ_H1, label = label),
            colour = "orange3",
            data = data.frame(
              M_IQ_H1 = 101.5,
              d_M_IQ_H1 = 0.04,
              label = " 二类错误")
```

```r
            ) +
    geom_text(aes(x = M_IQ_H1, y = d_M_IQ_H1, label = label),
              color = "orange3",
              data = data.frame(
              M_IQ_H1 = M_IQ_SAMPLE,
              d_M_IQ_H1 = 0.29,
              label = "H1 抽样分布")
            ) +
    scale_x_continuous(
      name = "M_IQ",
      breaks = c(90, 92, 94, 96, 200 - M_IQ_SAMPLE, LLCI_H0, 98, 100, 102, ULCI_H0, M_IQ_SAM
      guide = guide_axis(angle = 45, n.dodge = 2),
      minor_breaks = NULL
    )
H0H1plot
```




前面进行统计检验时，我们以双尾 $p < 0.05$ 为统计依据。如果 H0 确实是真的，我们以 $p < 0.05$ 为统计依据进行统计推断犯错的概率为 0.05，这种错误是一类错误 alpha。

如果 H0 确实是假的，而 H1 确实为真的呢？按照双尾 $p < 0.05$ 的统计标准，当 $p > 0.05$ 时，即 $97.06 < \text{M\_IQ} < 102.94$ 时（图中蓝色部分），我们会错误地得出结论：H0 为真，H1 为假。我们称这种错误为**二类错误（Type 2 error）**。我们犯这种错误的概率是多少呢？我们假定 H1 为真，所以此时应以 H1 的抽样分布为根据计算犯错概率，如上图所示，当 $97.06 < \text{M\_IQ} < 102.94$ 时我们会犯错，所以我们犯

错的概率为 H0 的 95%CI[97.06, 102.944] 与 H1 的概率密度曲线包围的黄色区域所代表的累积概率，我们可以使用 pnorm() 函数计算出二类错误的概率：

```
beta <- pnorm(q = qnorm(0.975, 100, sigma/sqrt(n)),
              M_IQ_SAMPLE,
              sigma/sqrt(n)) - pnorm(q = qnorm(0.025,
                                               100,
                                               sigma/sqrt(n)),
                                     M_IQ_SAMPLE,
                                     sigma/sqrt(n))
beta
## [1] 0.2991957
```

注意，H0 抽样分布与 H1 抽样分布的横坐标都是从负无穷到正无穷，因此，上图中 H1 的黄色区域实际上与 H0 的左侧尾部也有重叠，在计算 $\beta$ 时，我们需要减掉这一部分重叠面积所对应的概率。

另外，$(1-\beta)$ 被称为统计检验力 power，它反映了能正确识别 H1 与 H0 差异的能力，即正确拒绝 H0 的能力。

```
power <- 1 - beta
power
## [1] 0.7008043
```

我们可以使用 pwr 包中的 pwr.norm.test() 函数计算上述 z 检验的统计检验力。pwr.norm.test() 函数需要输入参数 Cohen's $d$，完整的计算如下：

```
library(pwr)
# 计算效应量 Cohen's d
d <- (M_IQ_SAMPLE - 100)/15
# 计算统计检验力
pwr.norm.test(d = d, n = n, alternative = "two.sided")
##
##      Mean power calculation for normal distribution with known variance
##
##               d = 0.2486667
##               n = 100
##       sig.level = 0.05
##           power = 0.7008043
##     alternative = two.sided
```

可见，pwr.norm.test() 的计算结果与前文计算的结果一致。

统计上一般要求统计检验力应当大于等于 0.80。提高统计检验力的方法有二，一是增大 H0 抽样分布与 H1 抽样分布之间的距离，即，增大效应量，二是降低 H0 抽样分布与 H1 抽样分布的离散程度，即，使这两个分布变得更加高瘦。如何降低 H0 抽样分布与 H1 抽样分布的离散程度呢？增大样本量。

你可以更改本文 R 代码一开始所设定的三个参数：`miuH1`、`sigma` 和 `n`，继而观察这三个参数的变化对 $p$ 值、beta、power 的影响。你也可以通过shiny动态演示 `miuH1`、`sigma` 和 `n` 对 $p$ 值、$\beta$、power 的影响——在 R 中运行下面的代码：

```r
# 检查并安装 shiny
if (!requireNamespace("shiny", quietly = TRUE)) {
  install.packages("shiny", dependencies = TRUE)
  message("Shiny 包安装成功！")

} else {
  message("Shiny 包已安装。")
}

# 加载 shiny 包，运行 H0H1plotShiny APP
library(shiny)
runUrl("https://www.psych.pub/files/H0H1plotShiny.zip")
```