



# 假设检验

## 极简介绍

高中统计学·第玖课

何濯羽

2024年2月23日



01

|| 假设检验的基本步骤

02

|| 卡方独立性检验

03

|| 不拒绝 $\neq$ 接受

04

|| 鸣谢



# 假设检验的基本步骤

## 步骤①：提出假设

假设我们正在研究总体中某一参数  $\theta$  的取值。我们希望  $\theta > \theta_0$ ，于是，我们可以构建如下假设：

$$H_0: \theta = \theta_0 \quad \text{和} \quad H_1: \theta > \theta_0$$

零假设

备择假设

**例** 假设随机变量  $X$ （代表学习时长）与  $Y$ （学习成绩）满足  $Y = 30 + \beta X + e$  的线性关系，其中  $\beta$  未知。我们希望  $\beta > 0$ （即学习成绩与学习时间是正相关的），于是，我们可以构建如下假设：

$$H_0: \beta = 0 \quad \text{和} \quad H_1: \beta > 0$$



## 步骤②：选择犯下 I 类错误的概率

	不拒绝 $H_0$	拒绝 $H_0$
$H_0$ 为真	正确的决定	I 类错误
$H_0$ 为假	II 类错误	正确的决定

犯下 I 类错误是概率被称为假设检验的**水准** (size)：

$$\text{size} = \Pr(\text{拒绝 } H_0 | H_0 \text{ 为真})$$

不犯 II 类错误的概率被称为假设检验的**效能** (power)：

$$\text{power} = 1 - \Pr(\text{不拒绝 } H_0 | H_0 \text{ 为假})$$

**注**

在同一假设检验中，犯 I 类错误的概率与犯 II 类错误的概率是此消彼长的关系，即我们无法同时降低犯两类错误的概率。针对此现象，研究员们通常采取的措施是：在可接受的范围内，选择一个最大的犯下 I 类错误的概率（记作  $\alpha$ ，被称为“**显著性水平**”； $1 - \alpha$  被称为“**置信水平**”），然后再尽可能地降低犯下 II 类错误的概率——即先选择一个较小的 size，然后尽可能地提高 power。

### 步骤③：构建统计量

由于我们无法直接观测总体，所以参数  $\theta$  的真实值不可知。我们可以观测的是  $\theta$  的估计值。

根据假设，我们构建一个**检验统计量** (test statistic)，记作  $T$ 。

构建  $T$  时必须满足两个条件：

- $T$  是一个以参数估计量  $\hat{\theta}$  为自变量的函数。
- 在  $\mathbb{H}_0$  为真且样本量趋向无穷大 ( $n \rightarrow \infty$ ) 时， $T$  服从特殊的概率分布  $\xi$  (例如正态分布)。
- 当  $\hat{\theta}$  的值与  $\mathbb{H}_0$  中  $\theta$  的假设值  $\theta_0$  相差较大时， $T$  的 (绝对) 值较大。

例

在  $Y = 30 + \beta X + e$  的线性模型中，我们的零假设为  $\mathbb{H}_0: \beta = 0$ 。研究员们通常会构建如下的检验统计量：

$$T = \frac{\hat{\beta} - 0}{s(\hat{\beta})}$$

其中， $s(\hat{\beta})$  是估计量  $\hat{\beta}$  的条件标准差， $s(\hat{\beta}) = \sqrt{D(\hat{\beta}|X)}$ ，常被称为参数估计的标准误 (standard error)。

这一统计量服从自由度为  $n - 1$  的学生  $t$  概率分布 (student t distribution)， $T \sim t(n - 1)$ 。

## 步骤④：计算临界值

计算临界值 (critical value) , 记作  $c$ 。该临界值必须满足  $1 - G(c) = \alpha$ , 其中,  $G(\cdot)$  是概率分布  $\xi$  的累积分布函数 (CDF) ,  $\alpha$  是我们在步骤②中选定的显著性水平 (即最大的可接受的犯下 I 类错误的概率) 。假设  $G(\cdot)$  是一个可逆函数, 则

$$c = G^{-1}(1 - \alpha)$$

注

在 Excel 中, 大多数特别的概率分布均有相应的函数可以计算它们的临界值。在选定函数后, 我们只需要为 Excel 再提供两个信息:  $\alpha$  的值和概率分布的自由度。

例如, 当我们选定  $\alpha = 0.05$  且检验统计量  $T$  服从  $t(50)$  时, 我们在 Excel 中输入的函数为

**T.INV(1-0.05, 50)**

返回结果为  $c = 1.676$  (保留3位小数) 。

## 步骤⑤：比较数值，做出决定

将参数估计值（和相关的样本数据）代入检验统计量  $T$ ，我们得到了检验统计值。将检验统计值与临界值相比较，做出决定：

- 如果  $T > c$ ，我们**可以拒绝**零假设  $\mathbb{H}_0$ ；
- 如果  $T \leq c$ ，我们**不可以拒绝**零假设  $\mathbb{H}_0$ 。

例

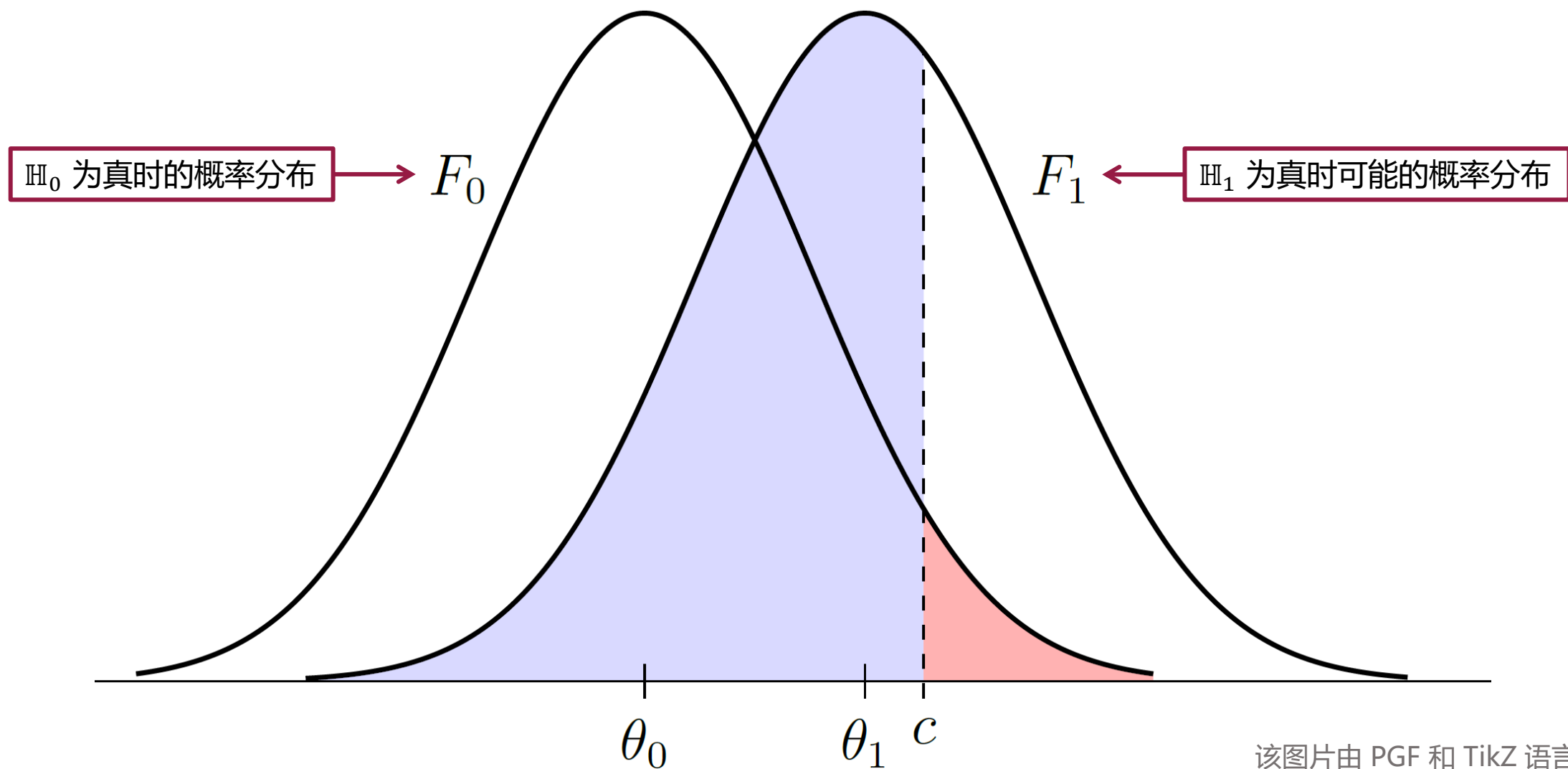
在学习成绩与学习时长的线性关系研究中，假设我们拥有样本量为 51 的样本数据，参数  $\beta$  的估计值为  $\hat{\beta} = 16$ ，参数估计量的标准误为  $s(\hat{\beta}) = 4$ ，则检验统计值为

$$T = \frac{\hat{\beta} - 0}{s(\hat{\beta})} = \frac{16 - 0}{4} = 4 > 1.676$$

因此，我们可以拒绝零假设  $\mathbb{H}_0: \beta = 0$ ，即“根据假设研究提供的证据，我们可以认为学习成绩与学习时长成正相关的关系”。



图示：单尾检验





# 卡方独立性检验



## 离散型随机变量的统计独立

当离散型随机变量  $X$  和  $Y$  的**联合概率** (joint probability) 等于它们**边际概率** (marginal probability) 之积时, 我们称  $X$  和  $Y$  相互 **(统计) 独立** (statistical independent) 。

$$\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

现实中, 我们无法直接观测总体, 即我们无法得知真实的联合概率  $\Pr(X = x, Y = y)$  和边际概率  $\Pr(X = x)$ 、 $\Pr(Y = y)$ 。因此, 我们选择通过抽样获得样本数据, 然后进行假设检验, 以寻找证明  $X$  与  $Y$  相互独立或不相互独立的证据。

## 例：爆米花与影片类型

假设我们从一个电影院收集了400名消费者的消费记录。我们可知每一名消费者的观看电影的类型（记作  $M$ ）以及他/她是否购买了爆米花（记作  $S$ ）。

- $M = 0$  代表“悲剧”， $M = 1$  代表“喜剧”；
- $S = 0$  代表“未购买爆米花”， $S = 1$  代表“购买爆米花”。

我们希望证明  $M$  与  $S$  不相互独立。



**步骤①：**提出假设。

$$\mathbb{H}_0: \Pr(M = m, S = s) = \Pr(M = m) \cdot \Pr(S = s) \quad \forall m \in \{0, 1\}, s \in \{0, 1\}$$

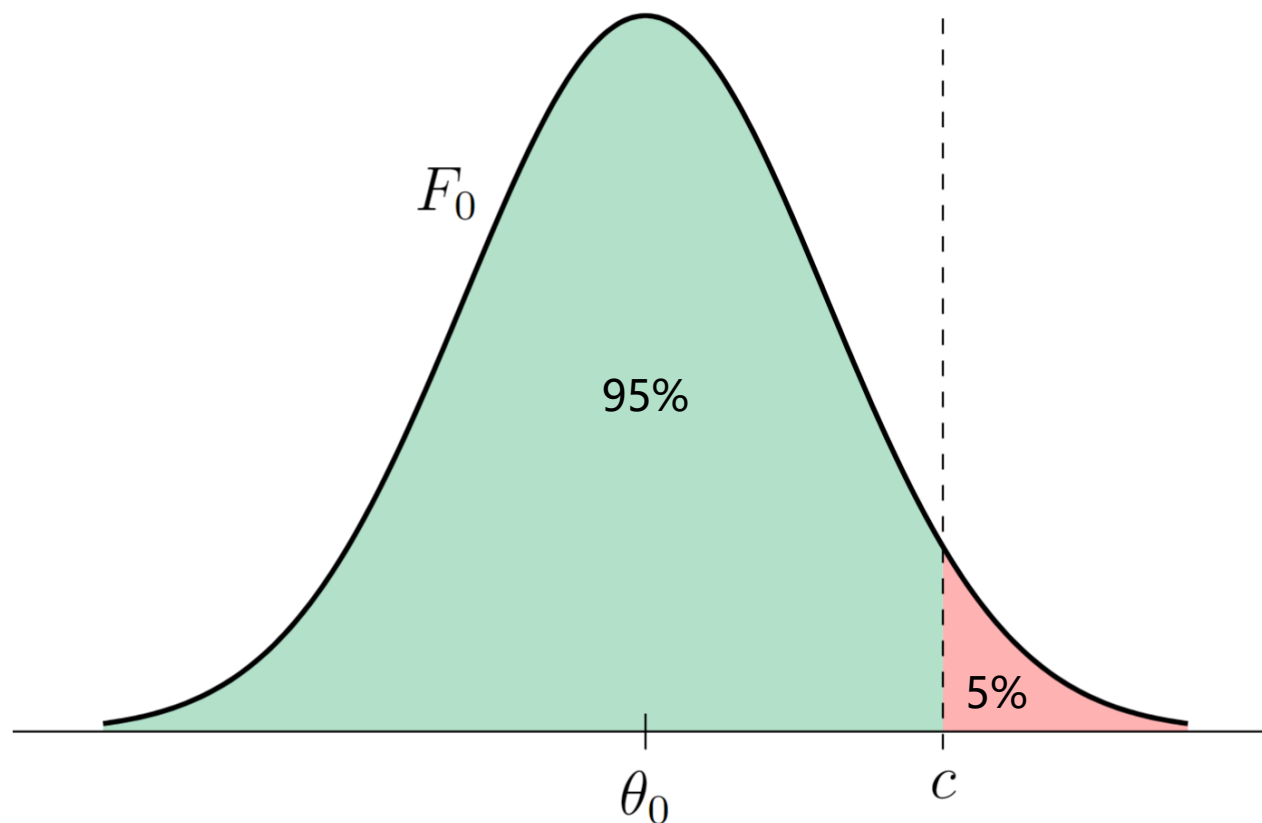
$$\mathbb{H}_1: \Pr(M = m, S = s) \neq \Pr(M = m) \cdot \Pr(S = s) \quad \exists m \in \{0, 1\}, s \in \{0, 1\}$$

## 例：爆米花与影片类型

**步骤②：**选择显著性水平。

高考中，显著性水平会由出题者给定。研究中，常用的显著性水平为  $\alpha = 0.1$ 、 $\alpha = 0.05$ 、 $\alpha = 0.01$ ，它们对应的置信水平依次为 90%、95%、99%。

此例中，我们选择  $\alpha = 0.05$ 。



## 例：爆米花与影片类型

步骤③：构建统计量。

假设我们收集的电影院消费者数据可被总结为如下所示的列联表（contingency table）：

	未买爆米花 ( $S = 0$ )	买过爆米花 ( $S = 1$ )	行总数
悲剧 ( $M = 0$ )	50	50	100
喜剧 ( $M = 1$ )	175	125	300
列总数	225	175	400

我们可以用频率去估计概率。例如，

$$\widehat{\Pr}(M = 0) = \frac{100}{400} = \frac{1}{4}$$

$$\widehat{\Pr}(M = 1) = \frac{300}{400} = \frac{3}{4}$$

$$\widehat{\Pr}(S = 0) = \frac{225}{400} = \frac{9}{16}$$

$$\widehat{\Pr}(S = 1) = \frac{175}{400} = \frac{7}{16}$$



## 例：爆米花与影片类型

步骤③：构建统计量。

	未买爆米花 ( $S = 0$ )	买过爆米花 ( $S = 1$ )	行总数
悲剧 ( $M = 0$ )	50	50	100
喜剧 ( $M = 1$ )	175	125	300
列总数	225	175	400

这样，在  $\mathbb{H}_0$  为真时，我们可以计算出每一种结果的期望频率（expected frequency rate）。

例如，

$$\hat{\pi}_{0,0} = \widehat{\Pr}(M = 0, S = 0) = \widehat{\Pr}(M = 0) \cdot \widehat{\Pr}(S = 0) = \frac{100}{400} \times \frac{225}{400} = \frac{9}{64}$$

接着，我们可以计算出每一种结果的期望频数。例如，

$$E_{0,0} = n \cdot \hat{\pi}_{0,0} = 400 \times \frac{9}{64} = \frac{225}{4} \approx 56$$

这一结果告诉我们：如果  $M$  和  $S$  相互独立，我们期望有 56 名消费者看了一场悲剧且没有购买爆米花。然而，真实数据显示只有 50 名消费者是如此的。

## 例：爆米花与影片类型

步骤③：构建统计量。

	未买爆米花 ( $S = 0$ )	买过爆米花 ( $S = 1$ )	行总数
悲剧 ( $M = 0$ )	50 56	50 44	100
喜剧 ( $M = 1$ )	175 169	125 131	300
列总数	225	175	400

上表中，黑色数字是样本中**观测到的频数**（observed count），而红色数字是  $H_0$  为真时我们**期望的频数**（expected count）。显然，现实与期望一定会有差距——假设检验将帮助我们检验这些差距的存在是由于误差的随机性，还是由于  $M$  与  $S$  并不相互独立。

## 例：爆米花与影片类型

步骤③：构建统计量。

构建独立性检验统计量一共分为四步：

- 1) 计算每一种结果相应的实际频数 ( $O_{i,j}$ ) 与期望频数 ( $E_{i,j}$ ) 的差；
- 2) 计算各差值的平方值（原理与方差的构建原理的相同）；
- 3) 将各平方值除以相应的期望频数；
- 4) 将以上所有的值加和，得到一个服从卡方分布（chi-squared distribution）的统计量：

$$T = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} = \sum_{i=1}^R \sum_{j=1}^C \frac{(n_{i,j} - n\hat{\pi}_{i,j})^2}{n\hat{\pi}_{i,j}} \sim \chi^2[(R-1)(C-1)]$$

其中，卡方分布的自由度为  $df = (R-1)(C-1)$ ， $R$  为第一个变量可能取值的个数（即列联表的**行数**）， $C$  为第二个变量可能取值的个数（即列联表的**列数**）。

在高考中，出题者往往会设置  $R = C = 2$ （即两个变量均只有 2 个可能取值），所以自由度为 1。

## 例：爆米花与影片类型

步骤④：计算临界值。

我们已选定  $\alpha = 0.05$ ，故临界值  $c$  为

$$c = G^{-1}(1 - \alpha) = G^{-1}(0.95)$$

其中， $G^{-1}(\cdot)$  为卡方分布的累积分布函数的反函数。在 Excel 中，我们可以使用 **CHISQ.INV( )** 或 **CHISQ.INV.RT( )** 函数计算临界值。例如，当自由度为  $df = 1$  时，

$\alpha$ 的值	0.1	0.05	0.01	0.005	0.001
左尾函数	CHISQ.INV(1-0.1,1)	CHISQ.INV(1-0.05,1)	CHISQ.INV(1-0.01,1)	CHISQ.INV(0.005,1)	CHISQ.INV(1-0.001,1)
右尾函数	CHISQ.INV.RT(0.1,1)	CHISQ.INV.RT(0.05,1)	CHISQ.INV.RT(0.01,1)	CHISQ.INV.RT(0.005,1)	CHISQ.INV.RT(0.001,1)
临界值	2.706	3.841	6.635	7.879	10.828

这就是人教版 2019 新教材 **表 8.3-4** 中各个数字的计算原理。

## 例：爆米花与影片类型

步骤⑤：比较数值，做出决定。

	未买爆米花 ( $S = 0$ )	买过爆米花 ( $S = 1$ )	行总数
悲剧 ( $M = 0$ )	50 56	50 44	100
喜剧 ( $M = 1$ )	175 169	125 131	300
列总数	225	175	400

根据以上样本数据，我们可以计算出统计量的假设值并将其与临界值  $c = \chi_{0.95}^2(1) \approx 3.841$  作比较：

$$T = \frac{(50 - 56)^2}{56} + \frac{(50 - 44)^2}{44} + \frac{(175 - 169)^2}{169} + \frac{(125 - 131)^2}{131} \approx 2.116 < 3.841$$

因此，根据显著性水平为 0.05（或置信水平为 95%）的卡方独立性检验（chi-squared test of independence）结果，我们无法拒绝零假设，即消费者在该电影院观看的电影类型与其是否购买爆米花“可能”是相互独立的。



不拒絕≠接受



谨防假设检验的滥用！



不拒绝零假设  $\neq$  必须接受零假设

在我们选定  $\alpha = 0.05$  的情况中， $T > c$  意味着我们的样本与  $\mathbb{H}_0$  为真时我们应该拥有的样本非常不一致。我们真实的样本只有 5% 的概率是抽取自  $\mathbb{H}_0$  为真时的概率分布  $F_0$ ，即犯下 I 类错误的概率为 5%。因此， $T > c$  是“拒绝  $\mathbb{H}_0$ ”的有力证据。

然而，当  $T \leq c$  时，我们没有证据足以拒绝  $\mathbb{H}_0$  —— 这并不等同于我们有证据足以接受  $\mathbb{H}_0$ 。谨记：当  $\alpha$  极小（即犯下 I 类错误的概率极小）时，那么犯下 II 类错误的概率  $\Pr(\text{接受 } \mathbb{H}_0 | \mathbb{H}_0 \text{ 为假})$  通常会非常大！



鸣谢

路漫漫其修远兮。  
感谢所有曾经在计量求索道路上给予我帮助的人们！

强迫症患者选择按初见的日期排序。

教授们



Jack Porter



Harold D. Chiang



Christopher R. Taber



Kenneth D. West



Mikkel Sølvsten



Brent Norwood



Myongjin Kim



Le Wang

求索者们



George Wu



Daniel Huo



Eric Zhou



JaeSeok Oh



Mieon Seong



Nikkie Gao



Chang Gao



Meng Yu


特别感谢



Xiaoxia Shi (不仅是教授, 还是Youtuber)



Bruce E. Hansen (拥有两本砖头的教授)



# 高考加油

高中统计学·第玖课

何濯羽

2024年2月23日