浙江大學 软件学院
School of Software Technology ,Zhejiang University

Part 6

假设检验
郎大为 J.D. Power

Outlines

假设检验

- 综述
- 两类错误
- 单边与双边
- p-value
- 常用的检验方法

综述

假设检验

- 假设(Hypothesis)是一个对一个或者多个总体的推断或者结论
- 假设检验(Hypothesis testing)是一个基于数据,用于决策的统计工具
- 为了将这个假设证明或者证伪,我们需要合适的知识来做出结论,也就是需要检验一部分总体
- 总的来说,假设检验的目的是:
  - 如何使用随机样本来判断
  - 随机样本能证实假设
  - 随机样本不能证实假设

假设检验

原假设与备择假设

- 假设检验由两部分构成
- H_0 null hypothesis
    - 原假设
- H_1 alternate hypothesis
    - 备择假设

假设检验

原假设与备择假设

- 我们想检验的是H_1 是否为真
  - 或者说"更像是"真的
- 假设检验有两种结论
  - 如果样本的更能证明H1
  - 如果样本不足以证明H1

假设检验

原假设与备择假设

- 我们想检验的是H_1 是否为真
  - 或者说"更像是"真的
- 假设检验有两种结论
  - 如果样本的更能证明H1
  - 如果样本不足以证明H1
- 拒绝H0, 接受H1
- 不拒绝H0

假设检验

原假设与备择假设

- **非常重要的内容!!!**
- 不拒绝H0不代表原假设H0为真
    - 不拒绝不等于接受
- 只代表并没有足够证据支持H1

假设检验

例子

- 法庭审判中
  - H0: 嫌疑人无罪
  - H1: 嫌疑人有罪
- 证据确凿:否认H0,接受H1,有罪
- 证据不足:不否认H0
- 无罪推定（presumption of innocence），又可称为无罪类推（与有罪类推相对应），简单地说是指任何人在未经依法判决有罪之前，应视其无罪

假设检验

例子

一家生产RAM芯片的公司声称该芯片的次品率为5%, 想检验的是:

- H0: 次品率$p = 5\%$
- H1: 次品率$p > 5\%$

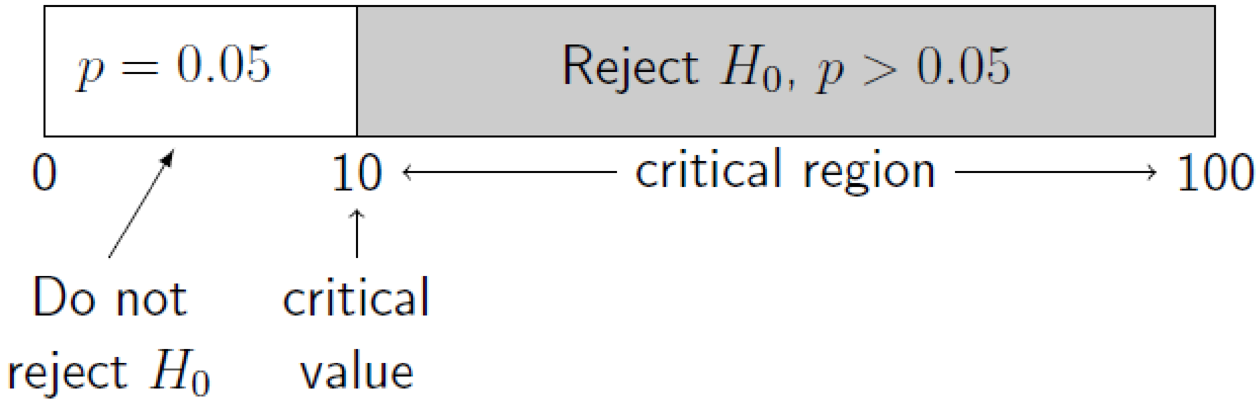
进行一个样本量为100的芯片抽样, 来进行假设检验

假设检验

例子

令X为100个样本中次品的数量,当 $x \geq 10$ 拒绝原假设

- X被称为检验统计量



假设检验

例子

- 为什么设置10作为检验的阈值?
- 每一次抽取是一个Bernoulli试验
- 整个过程是一个二项分布
- 如果H0为真
    - $n = 100$
    - $p = 0.05$
- 期望为$E(x) = 100 * 0.05 = 5$
- 因此10可以作为 $p > 5\%$ 的铁证~

两类错误

假设检验

两类错误

由于我们使用样本推断总体,所以犯错误是不可避免的,可能出现的错误有:

TRUTH	DECIDE	RESULT
H_0	H_0	Correctly accept null
H_0	H_a	Type I error
H_a	H_a	Correctly reject null
H_a	H_0	Type II error

假设检验

第一类错误

- 当H0为真的时候接受H1(拒绝H0)
    - α表示
    - '取伪'
- α 越小, 越不可能犯第一类错误
- 一般的假设检验会控制 α 在一个较小水平
- 比如0.05

假设检验

第一类错误

Case study continued

$$
\begin{aligned}
\alpha &= \Pr(\text{Type I error}) = \Pr(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \\
&= \Pr(X \geq 10 \text{ when } p = 0.05) \\
&= \sum_{x=10}^{100} b(x; n = 100, p = 0.05), \qquad \text{binomial distribution} \\
&= \sum_{x=10}^{100} \binom{100}{n} 0.05^x 0.95^{100-x} = 0.0282
\end{aligned}
$$

So, the level of significance is $\alpha = 0.0282$.

假设检验

第二类错误

- 当H1为真的时候不拒绝H0
    - β表示
    - '拒真'
- 除非知道很详细的备择假设,否则无法计算H1

假设检验

第二类错误

Case study continued

We cannot compute β for $H_1 : p > 0.05$ because the true p is unknown. However, we can compute it for testing $H_0 : p = 0.05$ against the alternative hypothesis that $H_1 : p = 0.1$, for instance.

$$\begin{aligned}\beta &= \Pr(\text{Type II error}) = \Pr(\text{reject } H_1 \text{ when } H_1 \text{ is true})\\ &= \Pr(X < 10 \text{ when } p = 0.1)\\ &= \sum_{x=0}^{9} b(x; n = 100, p = 0.1) = 0.4513\end{aligned}$$

假设检验

第二类错误

Case study continued

What is the probability of a type II error if $p = 0.15$?

$$\begin{aligned}\beta &= \Pr(\text{Type II error}) \\ &= \Pr(X < 10 \text{ when } p = 0.15) \\ &= \sum_{x=0}^{9} b(x; n = 100, p = 0.15) = 0.0551\end{aligned}$$

假设检验

两类错误的trade-off

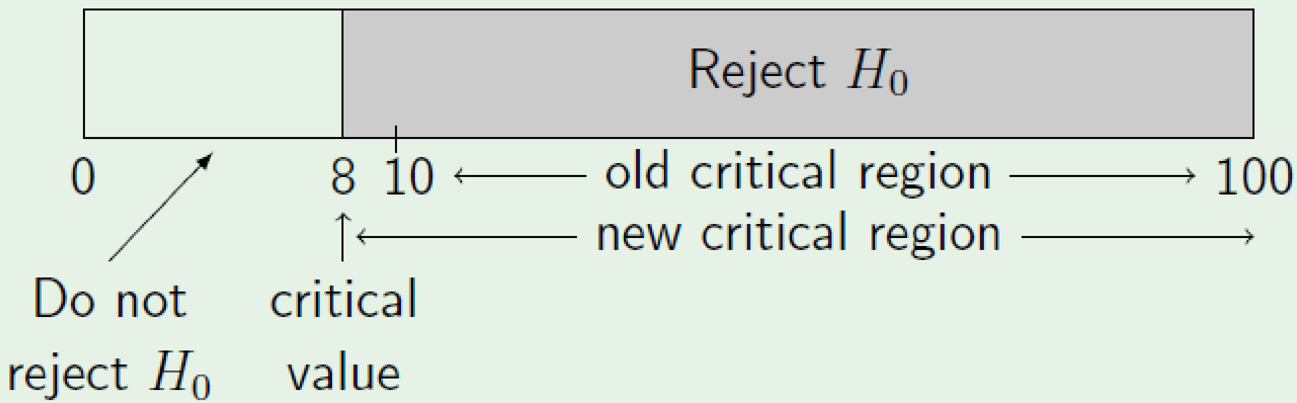
- 二者存在权衡关系
- 控制第二类错误的概率会放大拒绝域
  - 会使第一类错误的概率增加
  - β减少, α 增加
- 控制第一类错误会使第二类错误的概率增加
  - α减少, β 增加

假设检验

两类错误

Case study continued

Lets see what happens when we change the critical value from 10 to 8. That is, we reject H_0 if $X \geq 8$.



假设检验

两类错误

Case study continued

The new significance level is

$$\begin{aligned}\alpha &= \Pr(X \geq 8 \text{ when } p = 0.05) \\ &= \sum_{x=8}^{100} b(x; n = 100, p = 0.05) = 0.128.\end{aligned}$$

As expected, this is a large value than before (it was 0.0282).

假设检验

两类错误

- 扩大样本量是唯一减少两类错误的方式

Case study continued

Testing against the alternate hypothesis $H_1 : p = 0.1$,

$$\begin{aligned}\beta &= \Pr(X < 8 \text{ when } p = 0.1) \\ &= \sum_{x=0}^{7} b(x; n = 100, p = 0.1) = 0.206,\end{aligned}$$

which is lower than before.

Testing against the alternate hypothesis $H_1 : p = 0.15$,

$$\beta = \sum_{x=0}^{7} b(x; n = 100, p = 0.15) = 0.012,$$

again, lower than before.

假设检验

两类错误

Case study continued

Consider that now the sample size is $n = 150$ and the critical value is 12. Then, reject H_0 if $X \geq 12$, where X is now the number of defectives in the sample of 150 chips.

假设检验

两类错误

Case study continued

The significance level is

$$\begin{aligned}\alpha &= \Pr(X \geq 12 \text{ when } p = 0.05) \\ &= \sum_{x=12}^{150} b(x; n = 150, p = 0.05) = 0.074.\end{aligned}$$

Note that this value is lower than 0.128 for $n = 100$ and critical value of 8.

假设检验

两类错误

Case study continued

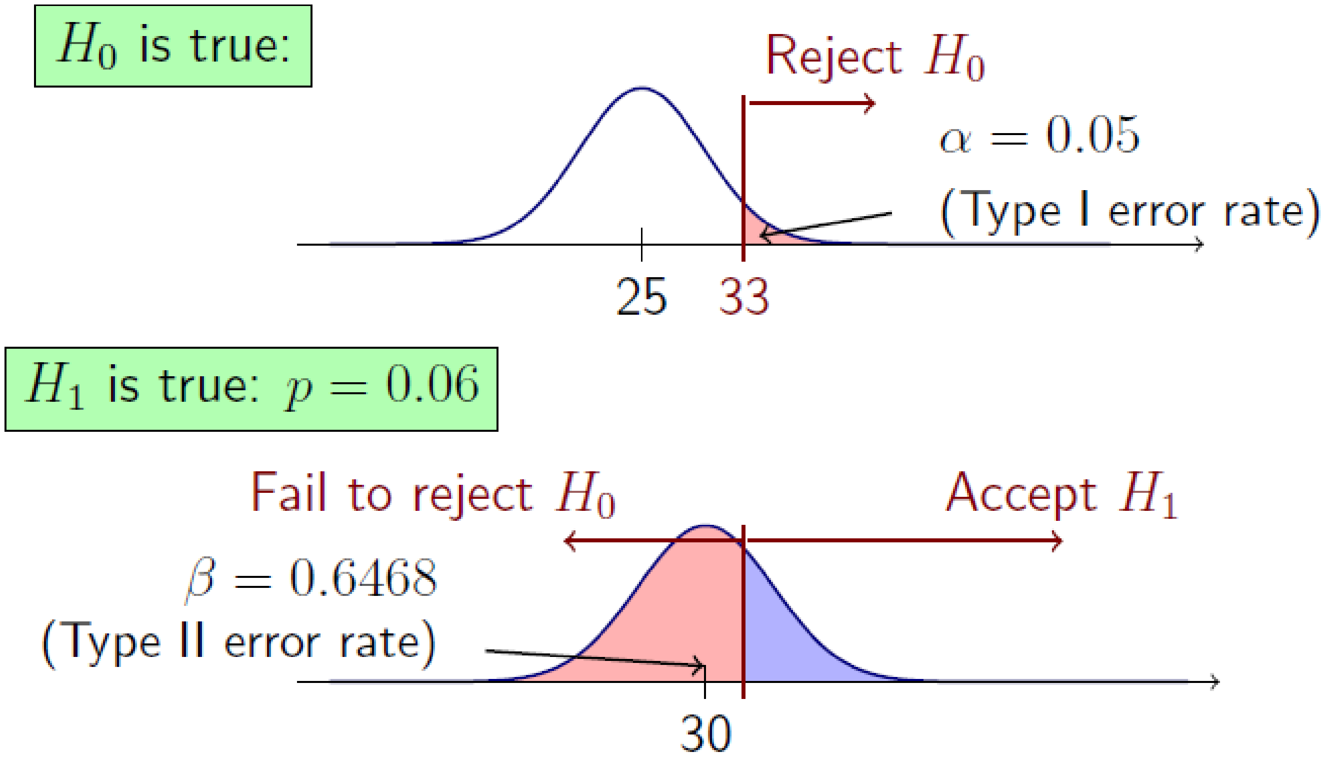
Testing against the alternate hypothesis $H_1 : p = 0.1$,

$$\begin{aligned}\beta &= \Pr(X < 12 \text{ when } p = 0.1) \\ &= \sum_{x=0}^{11} b(x; n = 150, p = 0.1) = 0.171,\end{aligned}$$

which is *also* lower than before (it was 0.206).

假设检验

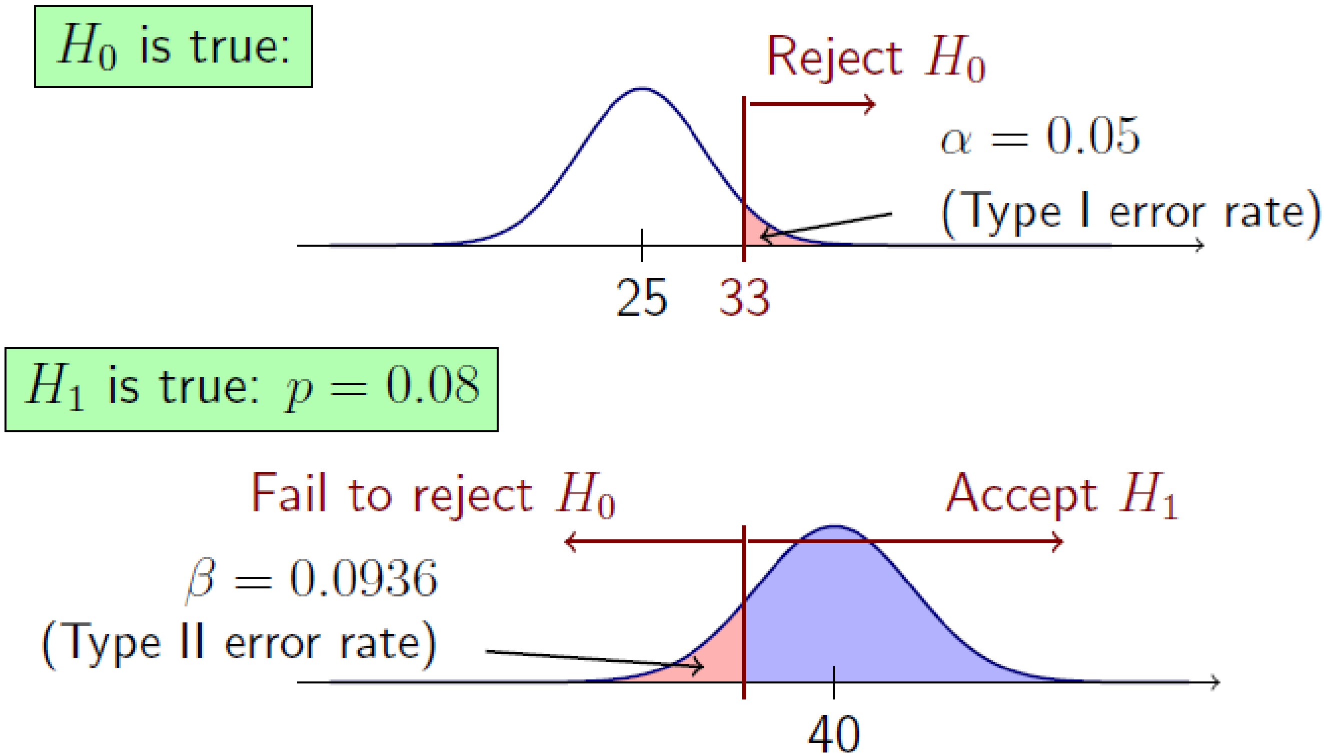
两类错误



H_0 is true:
Reject H_0
$\alpha = 0.05$
(Type I error rate)
25
33
H_1 is true: $p = 0.06$
Fail to reject H_0
Accept H_1
$\beta = 0.6468$
(Type II error rate)
30

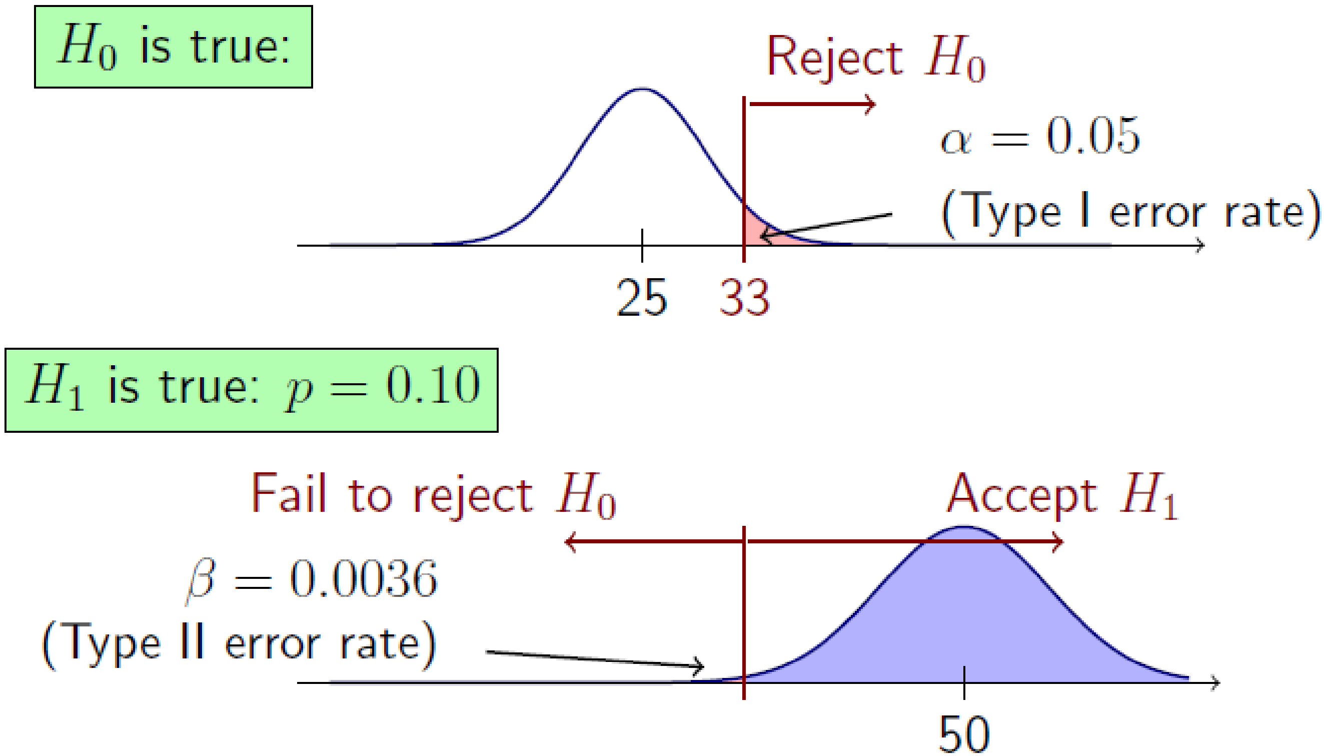
假设检验

两类错误


H_0 is true:
Reject H_0
$\alpha = 0.05$
(Type I error rate)
25
33
H_1 is true: $p = 0.08$
Fail to reject H_0
Accept H_1
$\beta = 0.0936$
(Type II error rate)
40

假设检验

两类错误



H_0 is true:
Reject H_0
$\alpha = 0.05$
(Type I error rate)
25
33
H_1 is true: $p = 0.10$
Fail to reject H_0
Accept H_1
$\beta = 0.0036$
(Type II error rate)
50

假设检验

Power

- 检验的势
- 当备择假设为真的时候拒绝H0的概率
- $$Power = 1 - \beta$$

假设检验

两类错误

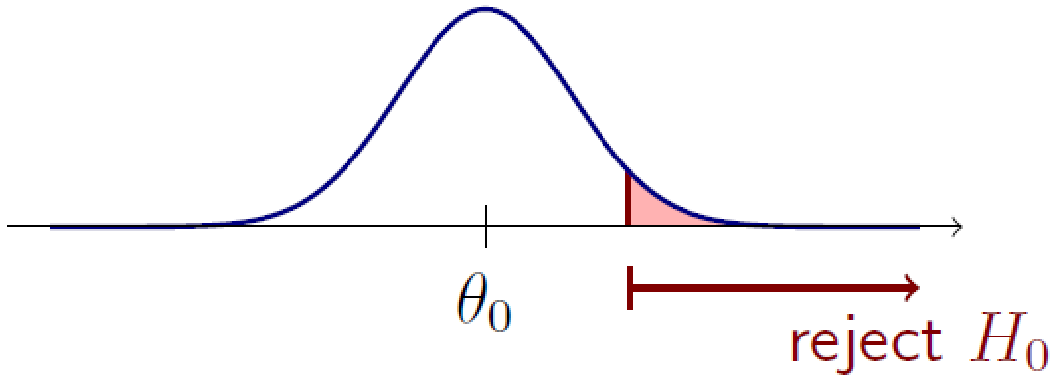
- α 和 β是相关联的, 一方增大会使另一方减少
- α可以被显著水平所控制, 比如0.05/0.1
- 增加样本量可以减少两类错误的概率
- 当H1与H0距离减少, β 的概率增加

单边与双边

单边与双边

单边检验

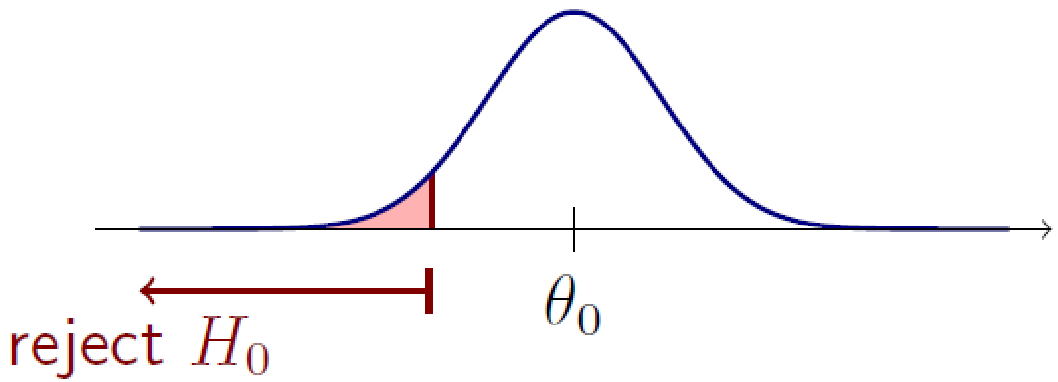
- 之前的例子,我们考虑的情况是:
  - H0: $\theta = \theta_0$
  - H1: $\theta > \theta_0$
- 这是一个经典的单边检验



单边与双边

单边检验

- H0: $\theta = \theta_0$
- H1: $\theta < \theta_0$



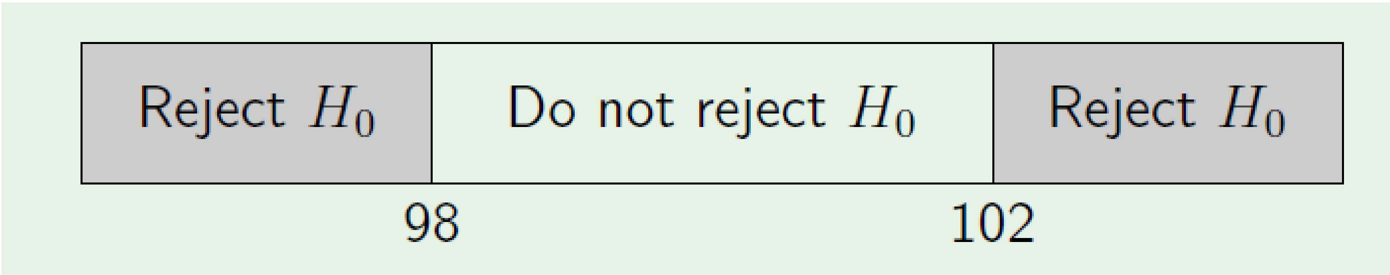
单边与双边

例子

某条生产线上生产的电阻X服从$N(100, 64)$, 假设检验为:

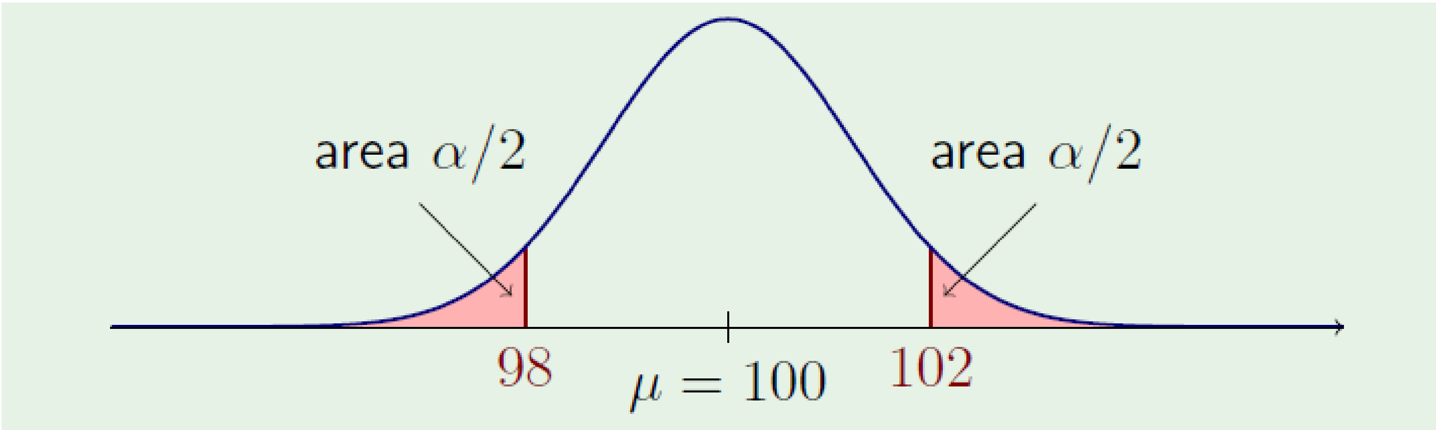
- H0: $\mu = 100$
- H1: $\mu \neq 100$

当样本量为100的时候:



单边与双边

例子

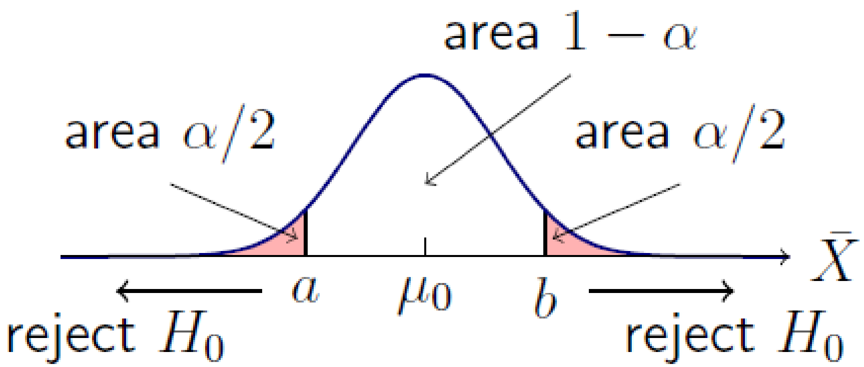


$\bar{(X)}$ 服从正态分布,均值为 μ , 标准差为 $\sigma/\sqrt{n}$

单边与双边

例子

$$\Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$



单边与双边

Z检验

- Z检验是检验一个正态总体中: $H_0 : \mu = \mu_0$ 对备择假设
  - $H_1 : \mu < \mu_0$
  - $H_2 : \mu \neq \mu_0$
  - $H_3 : \mu > \mu_0$
- 检验统计量的计算 $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- 拒绝域
  - $TS \leq -Z_{1-\alpha}$
  - $|TS| \geq Z_{1-\alpha/2}$
  - $TS \geq Z_{1-\alpha}$

单边与双边

t检验

- t检验是检验一个小样本的正态总体中: $H_0 : \mu = \mu_0$ 对备择假设
  - $H_1 : \mu < \mu_0$
  - $H_2 : \mu \neq \mu_0$
  - $H_3 : \mu > \mu_0$
- 检验统计量的计算 $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
- 拒绝域
  - $TS \leq -t_{1-\alpha}$
  - $|TS| \geq t_{1-\alpha/2}$
  - $TS \geq t_{1-\alpha}$

单边与双边

t检验

- 一个正态总体的小样本中
  - 服从t分布而不是正态分布
  - 自由度为 $n-1$
- 大样本下$t(n)$ 等价于$N(0,1)$

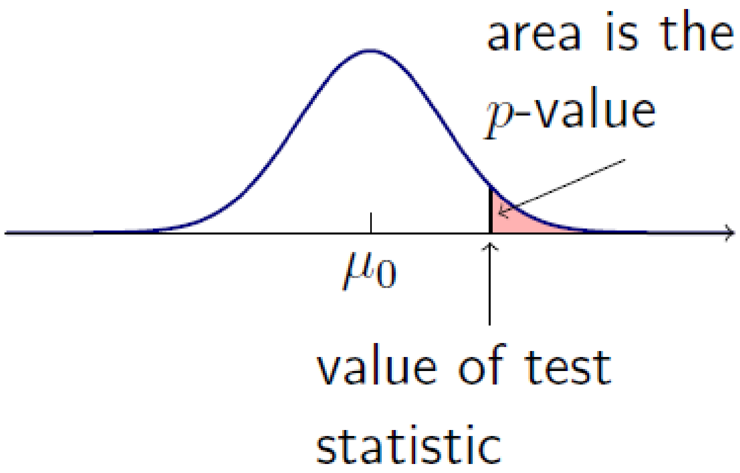
p-value

在一个检验中,拒绝H0(检验显著)的最小显著性水平.

p-value

另一个定义

- H0被拒绝时犯第?类错误的最?概率
  - 第二类,最小
  - 第二类,最大
  - 第一类,最小
  - 第一类,最大



p-value

例子

- 在一个假设检验中,$p = 0.09$
- 这个假设检验的结果是什么?
    - 拒绝H0, 接受H1?
    - 不拒绝H0?

p-value

例子

- 在一个假设检验中,$p = 0.09$
- 这个假设检验的结果是什么?
  - 拒绝H0, 接受H1?
  - 不拒绝H0?
- 分情况讨论
- 显著性水平为0.05
- 显著性水平为0.1

p-value

Example

A batch of 100 resistors have an average of 101.5 Ohms. Assuming a population standard deviation of 5 Ohms:

(a) Test whether the population mean is 100 Ohms at a level of significance 0.05.

(b) Compute the p-value.

p-value

Example continued

(a) $H_0 : \mu = 100, \quad H_1 : \mu \neq 100$

Test statistic is $\bar{X}$. Reject H_0 if

$$\bar{X} < 100 - z_{0.025}\frac{\sigma}{\sqrt{n}} = 100 - 1.96 \times \frac{5}{10} = 99.02$$

or

$$\bar{X} > 100 + z_{0.025}\frac{\sigma}{\sqrt{n}} = 100 + 1.96 \times \frac{5}{10} = 100.98$$

$\bar{X} = 101.5$ therefore, reject H_0.

p-value

Example continued

(b) The *observed* z-value is

$$Z = \frac{\bar{X} - 100}{\sigma/\sqrt{n}} = \frac{101.5 - 100}{5/10} = 3.$$

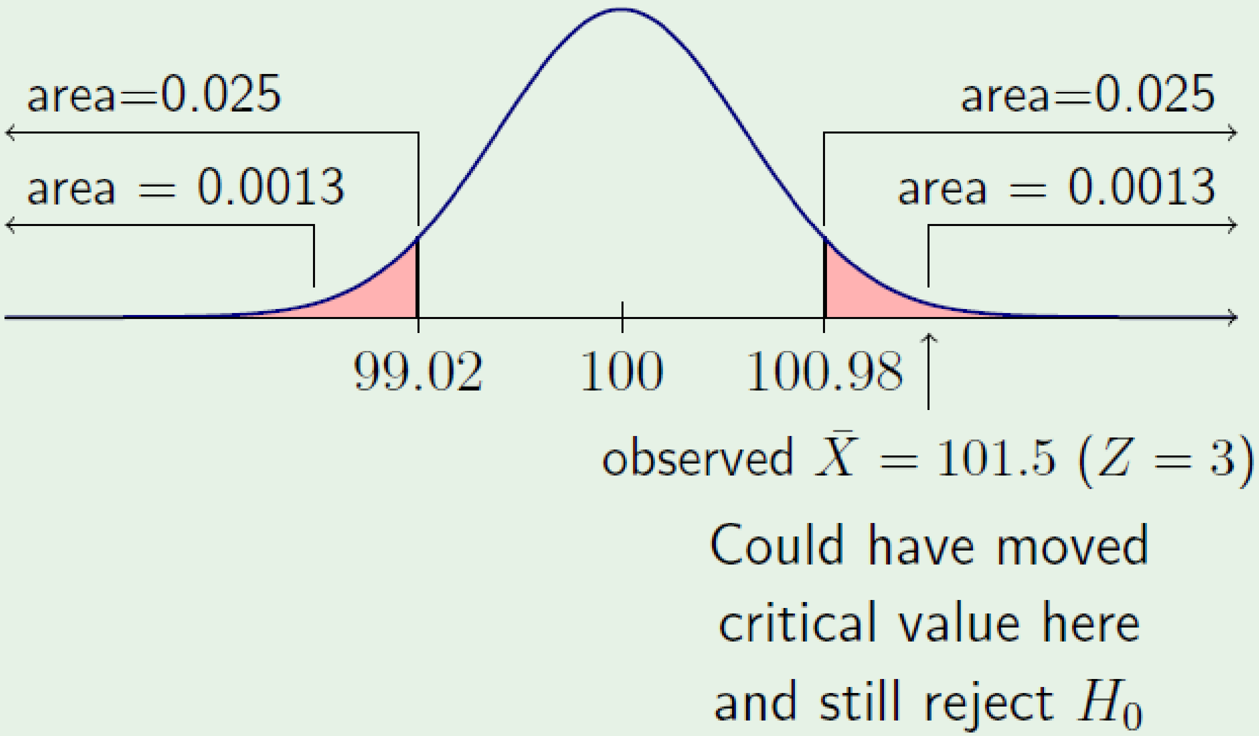
Then, the p-value is

$$p = 2\,\Pr(Z > 3) = 2 \times 0.0013 = 0.0026.$$

This means that H_0 could have been rejected at significance level $\alpha = 0.0026$ which is much stronger than rejecting it a 0.05.

p-value

Example continued



不同种类的假设检验

检验均值

- Z检验
- T检验
- 配对T 检验
  - H0: $\bar(x) = \mu$
  - H0: $\bar(x) = \bar(y)$
- 方差分析(ANOVA)
  - H0: $\bar(x) = \bar(y) = \bar(y)$
- 非参数检验(wilcox)

检验方差

- 卡方检验
  - H0: $\sigma^2 = s^2$
- F检验
  - H0: $\sigma_1^2 = k\sigma_2^2$

检验分布

- 列联表检验(卡方拟合优度检验)
    - Pearson's Chi-squared Test
    - H0: 两个频数表分布相似
- 正态性检验
    - Shapiro-Wilk Normality Test
    - H0: 样本服从正态分布

例子

- 检验两个城市的温度是否相同
  - (北京和广州)
- 方差分析前需要保证各样本方差相等
- 检验某个模型拟合是否优秀
- 检验线性回归的参数是否有效
- A/B test