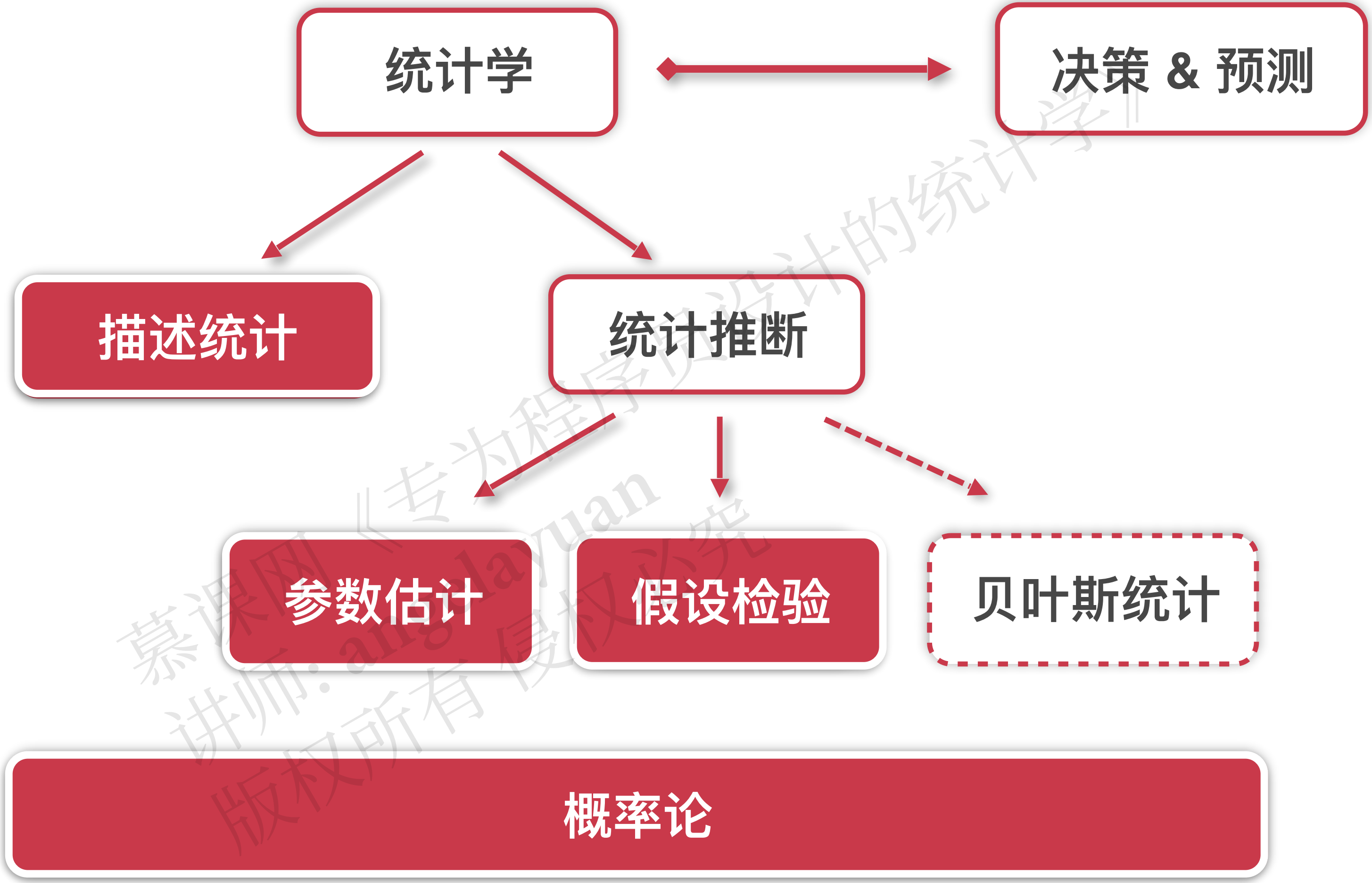

统计学
决策 & 预测
描述统计
统计推断
参数估计
假设检验
贝叶斯统计
概率论

什么是假设检验 (Hypothesis Testing)

频率学派推断 (Frequentist Inference)

H_0: 被告无罪

默认为真

H_A: 被告有罪

证据提供方

收集证据 (收集数据)

评估证据

(H_0为真时, 是否可能得到当前证据/数据?)

错误接受

是: 接受(fail to reject)H_0

错误拒绝

否: 拒绝(reject)H_0

假设检验

- 在总体的分布函数完全未知或只知其形式不知其参数的情况下，为了推断总体的某些未知特性，提出关于总体的假设
- 根据样本对所提出的假设做出是接受还是拒绝的决策
- 假设检验是做出这一决策的过程
- 我们不知道某一次决策是对是错, 但是遵守假设检验的流程, 我们至少知道犯错的频率

假设检验的要素

- 假设 (hypotheses)
  - 零假设 (null hypothesis; H_0): 默认为真

    30岁群体平均收入 $\mu = 1w$　　收入与性别无关 $\mu_F - \mu_M = 0$

  - 备择假设 (alternative hypothesis; H_A): 要检验的研究问题

    30岁群体平均收入 $\mu \neq 1w$　　收入与性别有关 $\mu_F - \mu_M \neq 0$

    $\neq, <, >$　　$\neq, <, >$

假设中涉及到的参数(μ)是总体参数

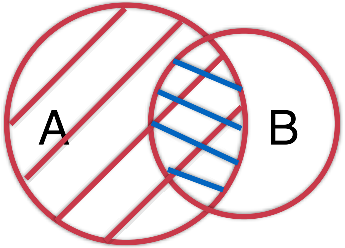
假设检验的要素

- **检验**
  - 在默认零假设为真的前提下，计算得到当前观测结果或更极端结果的概率

P(当前结果或更极端结果 | H_0为真)

设A, B是两个事件，且 $P(A) > 0$

则称 $P(B|A) = \dfrac{P(AB)}{P(A)}$ 为在事件A发生的条件下事件B发生的条件概率



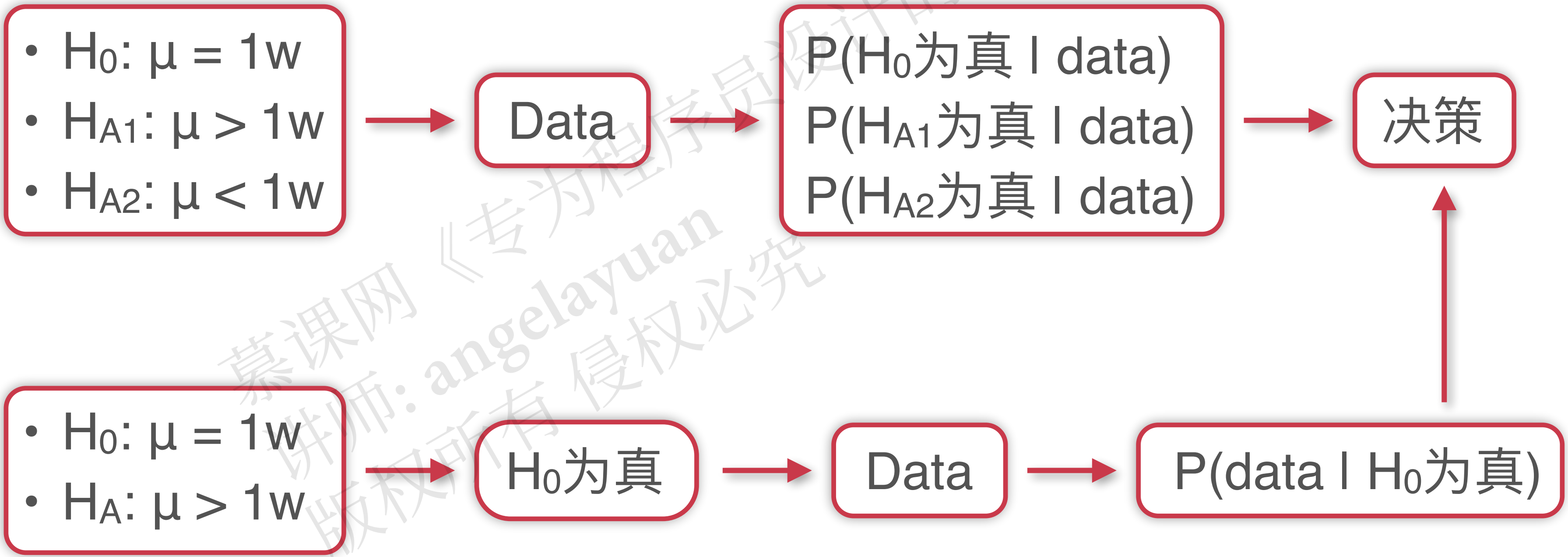
假设检验的要素

- **检验**
  - 如果数据无法提供令人信服的支持备择假设的证据，则接受零假设；如果数据提供了令人信服的支持备择假设的证据，则拒绝零假设

P(当前结果或更极端结果 | H_0为真)

- $p > \alpha$, 接受H_0
- $p \leq \alpha$, 拒绝H_0

假设检验: 频率论方法 vs 贝叶斯方法

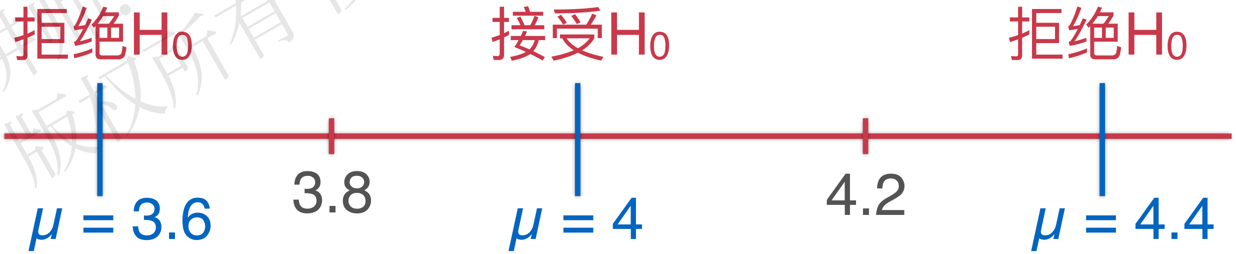


正态总体均值的假设检验

单个总体均值的检验

- 方差已知的情况: **置信区间检验法**
  - $H_0: \mu = 4$　　歌曲的时长服从正态分布，分布的均值μ为4
  - $H_A: \mu \neq 4$　　歌曲的时长服从正态分布，分布的均值μ不等于4

我们计算得到了一个置信水平为95%的歌曲平均时长的置信区间(3.8, 4.2)



单个总体均值的检验

- 方差已知的情况: **p值检验法**

歌曲的时长服从正态分布 $X \sim N(\mu, \sigma^2)$, $\sigma^2 = 1$; 手机里有25首歌曲，这25首歌曲的平均时长是4.5分钟

- H_0: $\mu = 4$
- H_A: $\mu > 4$

$\alpha = 0.05$

$\bar{X} = 4.5$

P(当前结果或更极端结果 | H_0为真)

$p > \alpha$, 接受H_0

$p \leq \alpha$, 拒绝H_0

α: 显著性水平

单个总体均值的检验

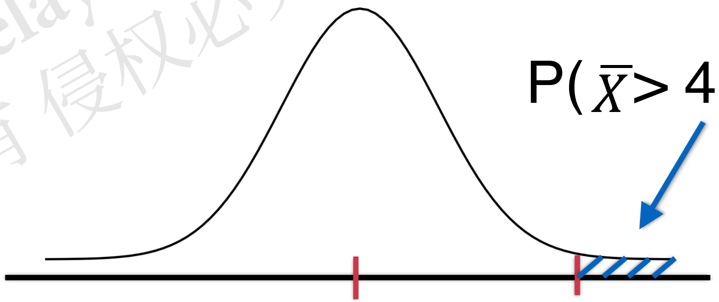
- 方差已知的情况: **p值检验法**
  - $H_0: \mu = 4$
  - $H_A: \mu > 4$

  $\alpha = 0.05$

  $\bar{X} = 4.5$

$P(\bar{X} > 4.5 \mid H_0: \mu = 4)$

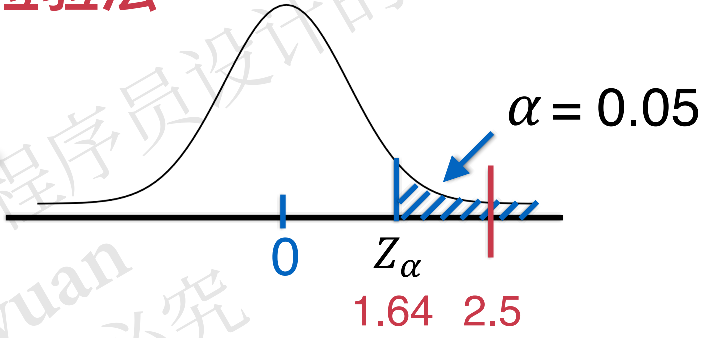
$\bar{X} \sim N(\mu, \sigma^2/n) \longrightarrow \bar{X} \sim N(4, 1/25)$

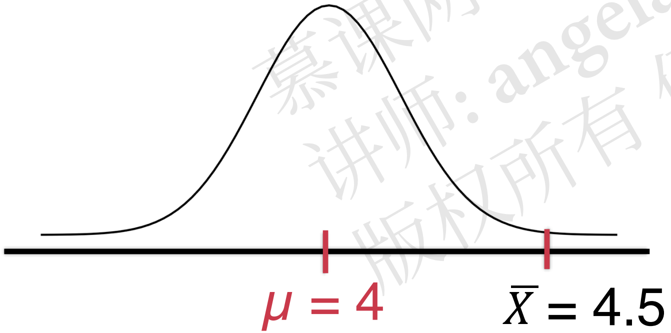


单个总体均值的检验

- 方差已知的情况: **p值检验法**

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$





$\bar{X} \longrightarrow Z$

$$\frac{4.5-4}{1/\sqrt{25}} = 2.5$$

$P(\bar{X} > 4.5 \mid H_0: \mu = 4)$

$P(Z > 2.5) = 0.006$ = p-value

p-value < α = 0.05 拒绝H_0

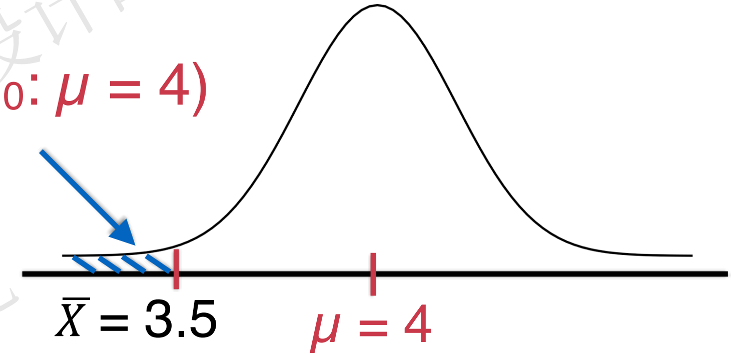
单个总体均值的检验

- 方差已知的情况: **p值检验法**
  - H_0: $\mu = 4$
  - H_A: $\mu < 4$

$\alpha = 0.05$

$\bar{X} = 3.5$

$P(\bar{X} < 3.5 \mid H_0: \mu = 4)$



$\bar{X} \longrightarrow Z$

$$\frac{3.5 - 4}{1/\sqrt{25}} = -2.5$$

$P(\bar{X} < 3.5 \mid H_0: \mu = 4)$

$P(Z < -2.5) = 0.006 =$ p-value

p-value $< \alpha = 0.05$ 拒绝H_0

单尾/边检验

one-tailed test

one-sided test

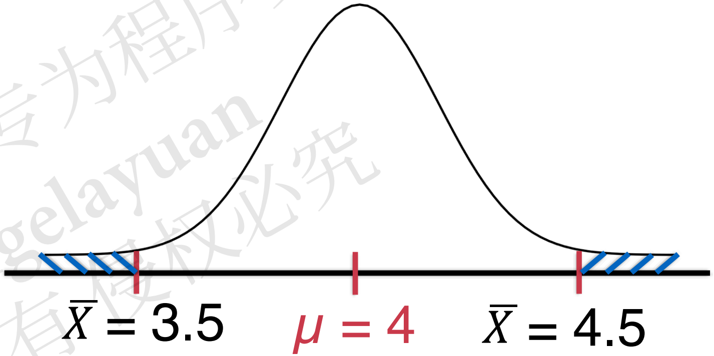
单个总体均值的检验

- 方差已知的情况: **p值检验法**
  - $H_0: \mu = 4$
  - $H_A: \mu \neq 4$

  $\alpha = 0.05$

  $\bar{X} = 4.5$

$P(\bar{X} > 4.5 \text{ or } \bar{X} < 3.5 \mid H_0: \mu = 4)$



Z检验

双尾/边检验

two-tailed test

two-sided test

$P(Z > 2.5 \text{ or } Z < -2.5) = 0.006 + 0.006 = 0.012$

p-value $< \alpha = 0.05$ 拒绝H_0

正态总体均值的假设检验

单个总体均值的检验

- **方差未知的情况: p值检验法**

歌曲的时长服从正态分布 $X \sim N(\mu, \sigma^2)$, σ^2未知; 手机里有25首歌曲， 这25首歌曲的平均时长是4.5分钟, 标准差为1分钟

- H_0: $\mu = 4$
- H_A: $\mu > 4$

$\alpha = 0.05$

$\bar{X} = 4.5$

$S = 1$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t\,(n-1)$$

$$\bar{X} \longrightarrow T \quad \frac{4.5 - 4}{1/\sqrt{25}} = 2.5$$

$$P(\bar{X} > 4.5 \mid H_0: \mu = 4) \longrightarrow P(T > 2.5)$$

单个总体均值的检验

- 方差未知的情况: **p值检验法**
  - $H_0: \mu = 4$
  - $H_A: \mu > 4$

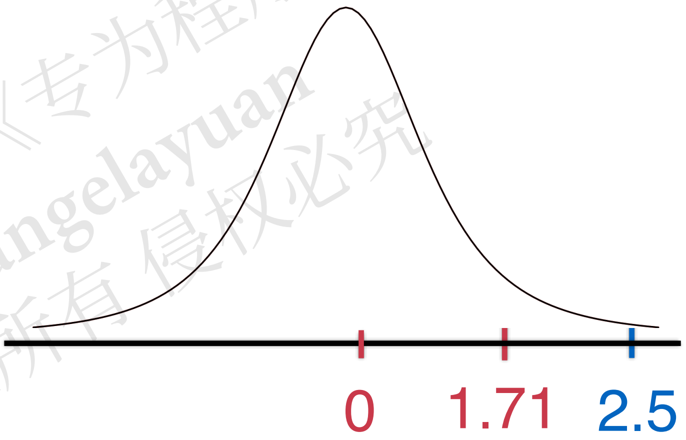
  $\alpha = 0.05$

  $\bar{X} = 4.5$

  $S = 1$

$P(\bar{X} > 4.5 \mid H_0: \mu = 4) \longrightarrow P(T > 2.5) = 0.01 < \alpha$

拒绝H_0



t检验

单样本t检验

One-sample t-test

单个总体均值的检验

- 方差未知的情况: **p值检验法**
  - $H_0: \mu = 4$
  - $H_A: \mu > 4$

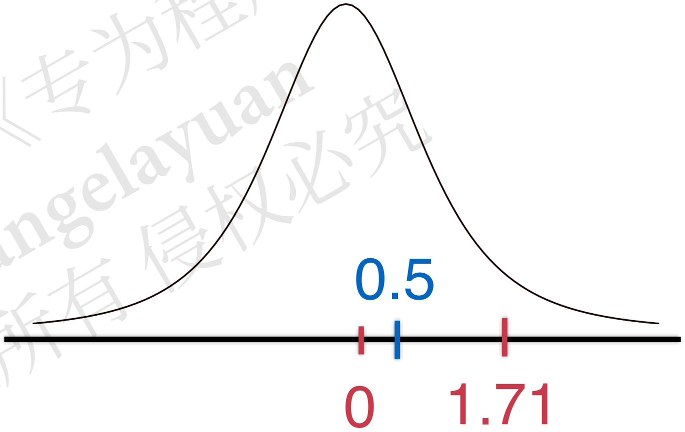
  $\alpha = 0.05$

  $\bar{X} = 4.1$

  $S = 1$

$P(\bar{X} > 4.1 \mid H_0: \mu = 4) \longrightarrow P(T > 0.5) = 0.31 > \alpha$

接受H_0



两个总体均值差的检验

- **方差已知的情况: p值检验法**

25左右人群的月收入服从正态分布$N(\mu_1,{\sigma_1}^2)$，35左右人群的月收入服从正态分布$N(\mu_2,{\sigma_2}^2)$, σ_1,σ_2 分别为2000和8000；我们记录了30名25岁和40名35岁个体的月收入。这30名25岁个体平均收入为16,000元，这40名35岁个体平均收入为25,000元

- H_0: $\mu_{25} - \mu_{35} = 0$
- H_A: $\mu_{25} - \mu_{35} \neq 0$

$\bar{X} = 16{,}000$ $\sigma_1 = 2000$ $n_1 = 30$

$\bar{Y} = 25{,}000$ $\sigma_2 = 8000$ $n_2 = 40$

$\alpha = 0.05$

两个总体均值差的检验

- 方差已知的情况: p值检验法
  - H_0: $\mu_{25} - \mu_{35} = 0$
  - H_A: $\mu_{25} - \mu_{35} \neq 0$

$\bar{X} = 16{,}000 \quad \sigma_1 = 2000 \quad n_1 = 30$

$\bar{Y} = 25{,}000 \quad \sigma_2 = 8000 \quad n_2 = 40$

$\alpha = 0.05$

P(当前结果或更极端结果 | H_0为真)

$P(\bar{X} - \bar{Y} < -9000 \text{ or } \bar{X} - \bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

两个总体均值差的检验

- 方差已知的情况: p值检验法

$P(\bar{X}-\bar{Y} < -9000 \text{ or } \bar{X}-\bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

$$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2})$$

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}} \sim N(0,1)$$

$$\bar{X}-\bar{Y} \longrightarrow Z$$

$$\frac{-9000-0}{\sqrt{\frac{2000^2}{30}+\frac{8000^2}{40}}} = -6.84$$

$$\frac{9000-0}{\sqrt{\frac{2000^2}{30}+\frac{8000^2}{40}}} = 6.84$$

两个总体均值差的检验

- 方差已知的情况: p值检验法

$P(\bar{X}-\bar{Y} < -9000$ or $\bar{X}-\bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

$\bar{X}-\bar{Y} \longrightarrow Z$

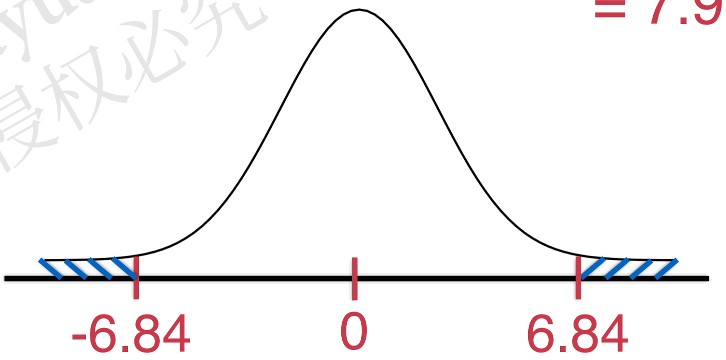
$$\frac{-9000-0}{\sqrt{\frac{2000^2}{30}+\frac{8000^2}{40}}} = -6.84$$

$$\frac{9000-0}{\sqrt{\frac{2000^2}{30}+\frac{8000^2}{40}}} = 6.84$$

$P(Z > 6.84$ or $Z < -6.84) = 3.96\text{e-}12 * 2$

$= 7.92\text{e-}12 < 0.05$

拒绝H_0



两个总体均值差的检验

- **方差未知且相等的情况: p值检验法**

25左右人群的月收入服从正态分布$N(\mu_1,\sigma_1{}^2)$，35左右人群的月收入服从正态分布$N(\mu_2,\sigma_2{}^2)$，σ_1,σ_2相等且未知；我们记录了30名25岁和40名35岁个体的月收入。这30名25岁个体平均收入为16,000元，标准差为2500元；这40名35岁个体平均收入为25,000元，标准差为7000元

- H_0: μ_{25} - μ_{35} = 0　　$\bar{X}$ = 16,000　　S_1 = 2500　　n_1 = 30
- H_A: μ_{25} - μ_{35} ≠ 0　　$\bar{Y}$ = 25,000　　S_2 = 7000　　n_2 = 40

α = 0.05

两个总体均值差的检验

- 方差未知且相等的情况: p值检验法
  - $H_0: \mu_{25} - \mu_{35} = 0$
  - $H_A: \mu_{25} - \mu_{35} \neq 0$

$\bar{X} = 16{,}000$　　$S_1 = 2500$　　$n_1 = 30$

$\bar{Y} = 25{,}000$　　$S_2 = 7000$　　$n_2 = 40$

$\alpha = 0.05$

P(当前结果或更极端结果 | H_0为真)

$P(\bar{X} - \bar{Y} < -9000 \text{ or } \bar{X} - \bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

两个总体均值差的检验

- 方差未知且相等的情况: p值检验法

P($\bar{X} - \bar{Y} < -9000$ or $\bar{X} - \bar{Y} > 9000$ | H_0: $\mu_{25} - \mu_{35} = 0$)

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

$$\bar{X} - \bar{Y} \longrightarrow T$$

$$\frac{\pm 9000 - 0}{5547 \times \sqrt{\frac{1}{30} + \frac{1}{40}}} = \pm 6.72$$

$$S_w = \sqrt{\frac{(n_1 - 1){S_1}^2 + (n_2 - 1){S_2}^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(30-1)\times 2500^2 + (40-1)\times 7000^2}{30 + 40 - 2}} = 5547$$

两个总体均值差的检验

- 方差未知且相等的情况: p值检验法

$P(\bar{X} - \bar{Y} < -9000 \text{ or } \bar{X} - \bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

$\bar{X} - \bar{Y} \longrightarrow T$

$$\frac{\pm 9000 - 0}{5547 \times \sqrt{\frac{1}{30} + \frac{1}{40}}} = \pm 6.72$$

t检验

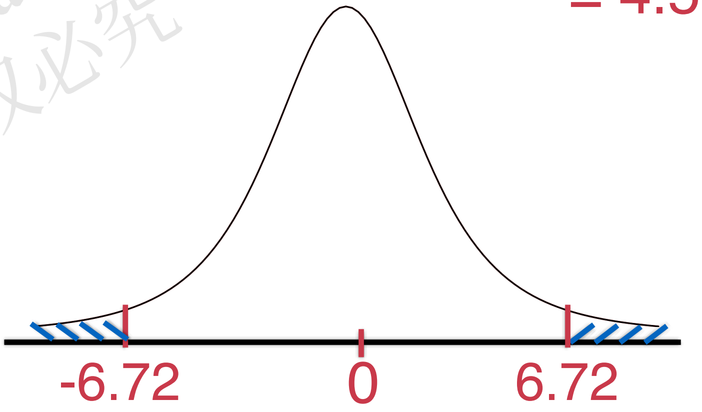
两个独立样本的t检验

Two independent sample t-test

$P(T > 6.72 \text{ or } T < -6.72) = 2.26\text{e-}9 * 2$

$= 4.52\text{e-}9 < 0.05$

拒绝H_0



两个总体均值差的检验

- **方差未知且不等的情况: p值检验法**

25左右人群的月收入服从正态分布$N(\mu_1, {\sigma_1}^2)$，35左右人群的月收入服从正态分布$N(\mu_2, {\sigma_2}^2)$，σ_1, σ_2 不等且未知；我们记录了30名25岁和40名35岁个体的月收入。这30名25岁个体平均收入为16,000元，标准差为2500元；这40名35岁个体平均收入为25,000元，标准差为7000元

- H_0: $\mu_{25} - \mu_{35} = 0$
- H_A: $\mu_{25} - \mu_{35} \neq 0$

$\bar{X} = 16{,}000$ $S_1 = 2500$ $n_1 = 30$

$\bar{Y} = 25{,}000$ $S_2 = 7000$ $n_2 = 40$

$\alpha = 0.05$

两个总体均值差的检验

- 方差未知且不等的情况: p值检验法
  - $H_0: \mu_{25} - \mu_{35} = 0$
  - $H_A: \mu_{25} - \mu_{35} \neq 0$

$\bar{X} = 16{,}000$ $S_1 = 2500$ $n_1 = 30$

$\bar{Y} = 25{,}000$ $S_2 = 7000$ $n_2 = 40$

$\alpha = 0.05$

P(当前结果或更极端结果 | H_0为真)

$P(\bar{X} - \bar{Y} < -9000 \text{ or } \bar{X} - \bar{Y} > 9000 \mid H_0: \mu_{25} - \mu_{35} = 0)$

两个总体均值差的检验

- 方差未知且不等的情况: p值检验法

P($\bar{X} - \bar{Y}$ < -9000 or $\bar{X} - \bar{Y}$ > 9000 | H0: μ_{25} - μ_{35} = 0)

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{S_1}^2}{n_1}+\frac{{S_2}^2}{n_2}}} \sim t$$

$$\bar{X}-\bar{Y} \longrightarrow T$$

$$\frac{\pm 9000 - 0}{\sqrt{\frac{2500^2}{30}+\frac{7000^2}{40}}} = \pm 7.52$$

$$df = \frac{\left(\frac{{S_1}^2}{n_1}+\frac{{S_2}^2}{n_2}\right)^2}{\frac{({S_1}^2/n_1)^2}{n_1-1}+\frac{({S_2}^2/n_2)^2}{n_2-1}} = \frac{\left(\frac{2500^2}{30}+\frac{7000^2}{40}\right)^2}{\frac{(2500^2/30)^2}{30-1}+\frac{(7000^2/40)^2}{40-1}} = 51.4$$

两个总体均值差的检验

- 方差未知且不等的情况: p值检验法

P($\bar{X} - \bar{Y} < -9000$ or $\bar{X} - \bar{Y} > 9000$ | H_0: $\mu_{25} - \mu_{35} = 0$)

$\bar{X} - \bar{Y} \longrightarrow T$

$$\frac{\pm 9000 - 0}{\sqrt{\frac{2500^2}{30} + \frac{7000^2}{40}}} = \pm 7.52$$

P($T > 7.52$ or $T < -7.52$)

= 3.93e-10 * 2

= 7.86e-10 < 0.05 拒绝H_0

t检验

两个独立样本的t检验

Two independent sample t-test

基于成对数据的检验

不满足互相独立

编号	数学成绩_训练前	数学成绩_训练后
1	70	80
2	85	91
3	95	92
4	90	89

基于成对数据的检验

一个变量/总体

编号	数学成绩_训练前	数学成绩_训练后	成绩差_(后-前)
1	70	80	10
2	85	91	6
3	95	92	-3
4	90	89	-1

基于成对数据的检验

- 参数: 所有参加训练的学生的两次成绩之差的平均值 $\mu_{\text{post-pre}}$
- 点估计: 样本包含的学生的两次成绩之差的平均值

- $H_0: \mu_{\text{post-pre}} = 0$　训练前后数学成绩没有差别
- $H_A: \mu_{\text{post-pre}} \neq 0$　训练前后数学成绩有差别
- $H_A: \mu_{\text{post-pre}} > 0$　训练后数学成绩有提高
- $H_A: \mu_{\text{post-pre}} < 0$　训练后数学成绩有下降

基于成对数据的检验

从所有参加训练的学生中抽取了50人，采集了训练前后的数学成绩；对每一个学生计算数学成绩之差(训练后-训练前)后，得到成绩之差的均值为10, 标准差为3；问题: 训练是否有效提高了数学成绩?

- H_0: $\mu_{\text{post-pre}} = 0$
- H_A: $\mu_{\text{post-pre}} > 0$

$\bar{X}_{post-pre} = 10$ $\quad n_{post-pre} = 50$

$S_{post-pre} = 3$ $\quad \alpha = 0.05$

基于成对数据的检验

- H_0: $\mu_{\text{post-pre}} = 0$
- H_A: $\mu_{\text{post-pre}} > 0$

$\bar{X}_{post-pre} = 10$ $n_{post-pre} = 50$

$S_{post-pre} = 3$ $\alpha = 0.05$

P(当前结果或更极端结果 | H_0为真)

$P(\bar{X}_{post-pre} > 10 \mid H_0: \mu_{\text{post-pre}} = 0)$

$$\bar{X}_{post-pre} \longrightarrow T$$

$$\frac{10-0}{3/\sqrt{50}} = 23.57$$

$P(T > 23.57) = 1.09\text{e-}28$

拒绝H_0

t检验

配对t检验

paired t-test

单个 vs 两个总体均值的检验

单个总体均值的检验
以及配对t检验

1个数值变量

编号	身高
1	165 cm
2	175 cm
3	170 cm
4	168 cm

编号	身高差
1	5 cm
2	10 cm
3	7 cm
4	6 cm

单个 vs 两个总体均值的检验

两个水平

两个总体均值的检验 ⟷ 1个数值变量和1个分类变量

Response variable: 身高

Explanatory variable: 性别

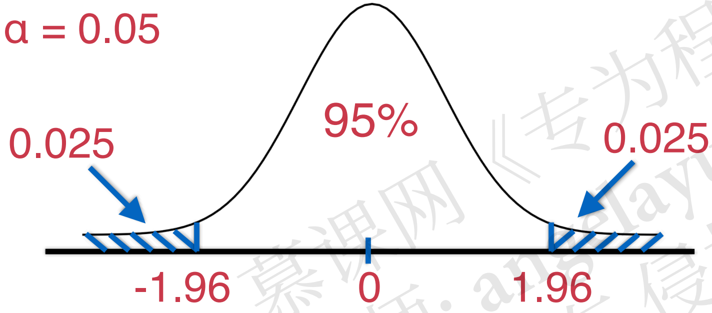
编号	身高	性别
1	165 cm	女
2	175 cm	男
3	170 cm	女
4	168 cm	男

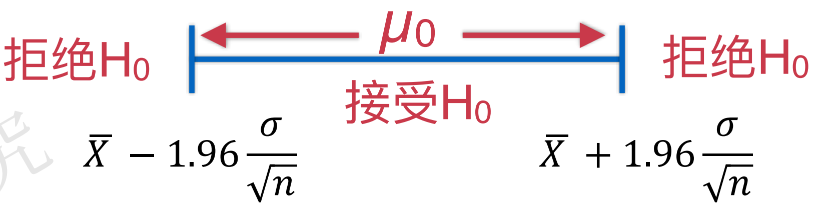
置信区间与假设检验的关系

置信区间与假设检验的关系

显著性水平为α的双尾假设检验 ⟷ 置信水平为1 - α的双侧置信区间

- $H_0: \mu = \mu_0$
- $H_A: \mu \neq \mu_0$





$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \longrightarrow \mu_0 = \bar{X} - Z\frac{\sigma}{\sqrt{n}}$$

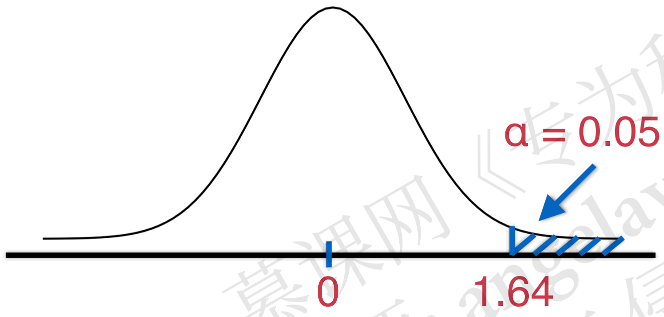
$$Z \geq 1.96,\quad \mu_0 \leq \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}$$

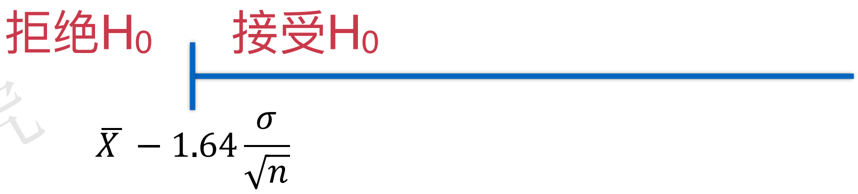
$$Z \leq -1.96,\quad \mu_0 \geq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

置信区间与假设检验的关系

显著性水平为α的单尾假设检验 ⟷ 置信水平为1 - α的单侧置信区间

- $H_0: \mu = \mu_0$
- $H_A: \mu > \mu_0$





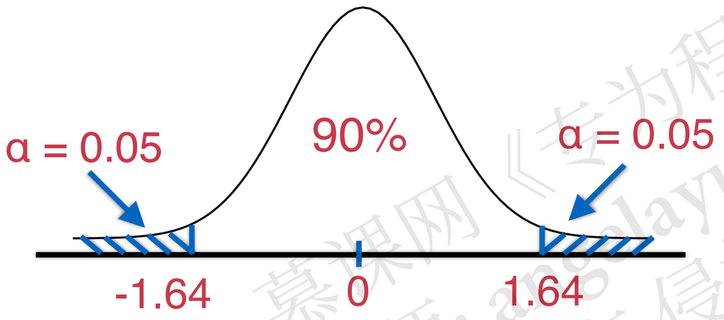
$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \longrightarrow \mu_0 = \bar{X} - Z\frac{\sigma}{\sqrt{n}} \longrightarrow Z \geq 1.64,\ \ \mu_0 \leq \bar{X} - 1.64\frac{\sigma}{\sqrt{n}}$$

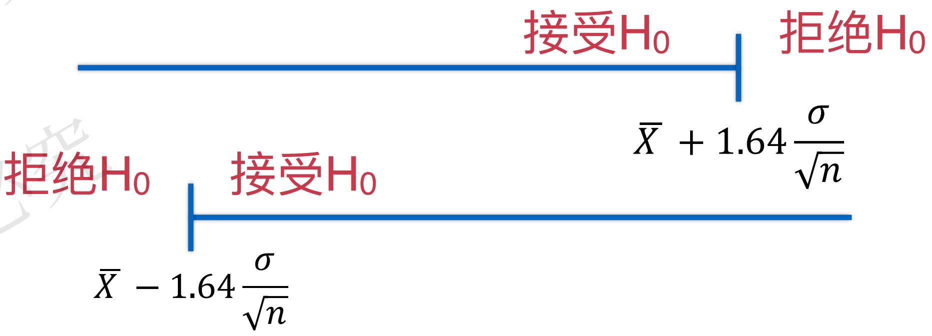
置信区间与假设检验的关系

显著性水平为α的单尾假设检验 ⟷ 置信水平为1 - α的单侧置信区间

置信水平为1 - 2α的双侧置信区间

- $H_0: \mu = \mu_0$
- $H_A: \mu < \mu_0$





$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \longrightarrow \mu_0 = \bar{X} - Z\frac{\sigma}{\sqrt{n}} \longrightarrow Z \le -1.64,\ \ \mu_0 \ge \bar{X} + 1.64\frac{\sigma}{\sqrt{n}}$$

置信区间与假设检验的关系

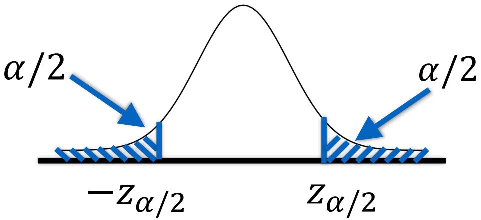
- 如果假设检验的结果是拒绝H_0，则置信区间应该不包含H_0中的参数值

- 如果假设检验的结果是接受H_0，则置信区间应该包含H_0中的参数值

编程实现正态总体均值的假设检验

单个总体

$$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$



$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t\,(n-1)$$

两个总体

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}}}\sim N(0,1)$$

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{{S_1}^2}{n_1}+\frac{{S_2}^2}{n_2}}}\sim t$$

$$\frac{\bar{X}_{diff}-\mu_{diff}}{S_{diff}/\sqrt{n}}\sim t\,(n-1)$$

配对t检验

正态总体方差的假设检验

单个总体方差的检验

25岁左右人群的月收入在3年前服从标准差为2(千元)的正态分布; 我们现收集了30名25岁个体的月收入，标准差为2.5千元。问能够推断现在25岁人群月收入的波动性与3年前有显著变化吗?

- $H_0: \sigma^2 = 4$
- $H_A: \sigma^2 \neq 4$

$S = 2.5, n = 30$

$\alpha = 0.05$

P(当前结果或更极端结果 | H_0为真)

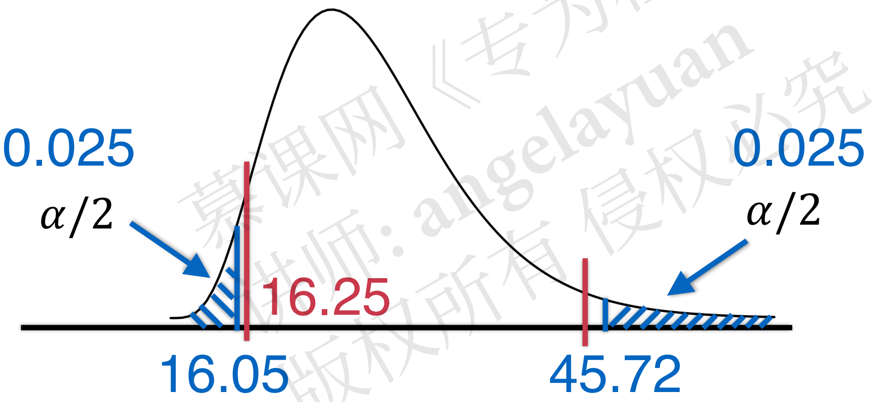
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$S^2 \longrightarrow \chi^2 \qquad \frac{(30-1)\times 2.5^2}{4} = 45.31$$

单个总体方差的检验

P(当前结果或更极端结果 | H_0为真)

$S^2 \longrightarrow \chi^2 \quad \frac{(30-1)\times 2.5^2}{4} = 45.31$



p-value = **2***P(χ^2> 45.31)

= 2*0.027 = 0.054

> 0.05

接受H_0

卡方检验

两个总体方差比的检验

我们现收集了30名25岁个体的月收入，标准差为2.5千元; 收集了40名35岁个体的月收入，标准差为4.2千元。问能够推断25岁人群月收入的波动小于35岁人群月收入的波动吗?

- H_0: $\frac{\sigma_{25}{}^2}{\sigma_{35}{}^2} = 1$
- H_A: $\frac{\sigma_{25}{}^2}{\sigma_{35}{}^2} < 1$

$S_1 = 2.5$ $n_1 = 30$

$S_2 = 4.2$ $n_2 = 40$

$\alpha = 0.05$

$$\frac{S_1{}^2/S_2{}^2}{\sigma_1{}^2/\sigma_2{}^2} \sim F\,(n_1 - 1, n_2 - 1)$$

$$\frac{S_1{}^2}{S_2{}^2} \longrightarrow F$$

两个总体方差比的检验

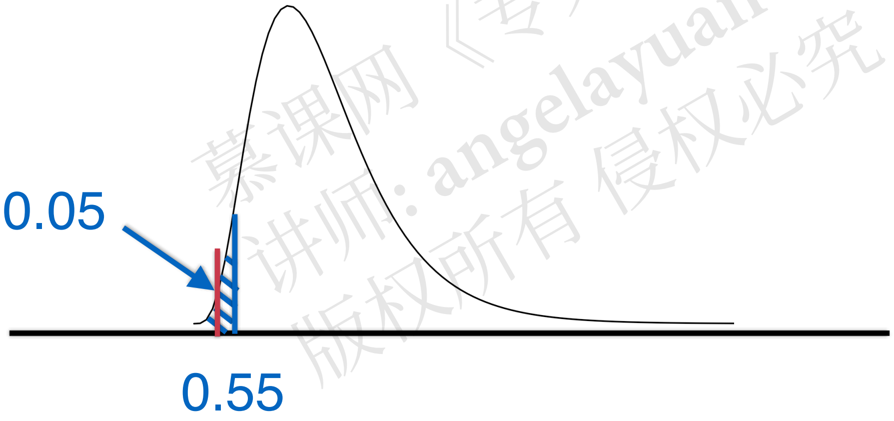
P(当前结果或更极端结果 | H_0为真)

$$\frac{{S_1}^2}{{S_2}^2} \longrightarrow F \qquad \frac{2.5^2/4.2^2}{1} = 0.35$$

p-value = P(F < 0.35) = 0.002

< 0.05

拒绝H_0



F检验

两个总体方差比的检验

- 检查是否满足方差相等这个前提条件

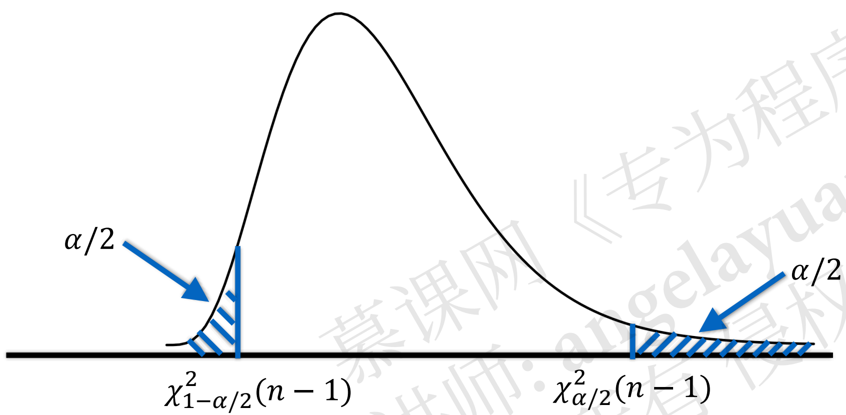
我们想检验25岁群体和35岁群体的平均月收入是否有显著差异，但是我们不知道这两个群体的方差是否相等(方差是否相等决定了应该使用哪一种t检验)，所以我们先对方差是否相等进行检验

两总体方差相等也称两总体具有方差齐性

编程实现正态总体方差的假设检验

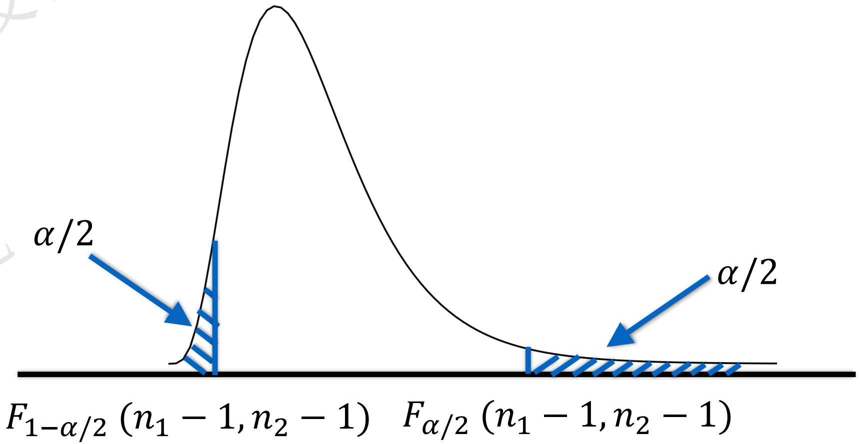
单个总体

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$



两个总体

$$\frac{{S_1}^2/{S_2}^2}{{\sigma_1}^2/{\sigma_2}^2} \sim F\,(n_1-1, n_2-1)$$



决策错误与统计功效

P(当前结果或更极端结果 | H_0为真)

→ $p > \alpha$, 接受H_0（结果不显著)

→ $p \le \alpha$, 拒绝H_0（结果显著)

		决策	
		接受H_0	拒绝H_0
真相	H_0为真	√	一类错误 type 1 error
	H_A为真	二类错误 type 2 error	√

P(reject H_0 | H_0 true) = α

P(accept H_0 | H_A true) = β

相互制约

H_0: 被告无罪

H_A: 被告有罪

- 决策错误
  - 被告是无辜的，但是法院判决被告有罪 → 1类错误
  - 被告不是无辜的，但是法院判决被告无罪 → 2类错误

哪类错误的后果更严重?

- 被告是无辜的，但是法院判决被告有罪 → 1类错误
- 被告不是无辜的，但是法院判决被告无罪 → 2类错误

哪类错误的后果更严重?

1类错误更严重，则选择低的显著性水平 (e.g. $\alpha = 0.01$)

2类错误更严重，则选择高的显著性水平 (e.g. $\alpha = 0.10$)

目标: 让犯1,2类错误的概率都尽量小

		决策	
		接受H_0	拒绝H_0
真相	H_0为真	**$1 - \alpha$**	$P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$
	H_A为真	$P(\text{accept } H_0 \mid H_A \text{ true}) = \beta$	**$1 - \beta$**

2类错误 ↓ = 统计功效 ↑

统计功效 (statistical **power**)

统计功效

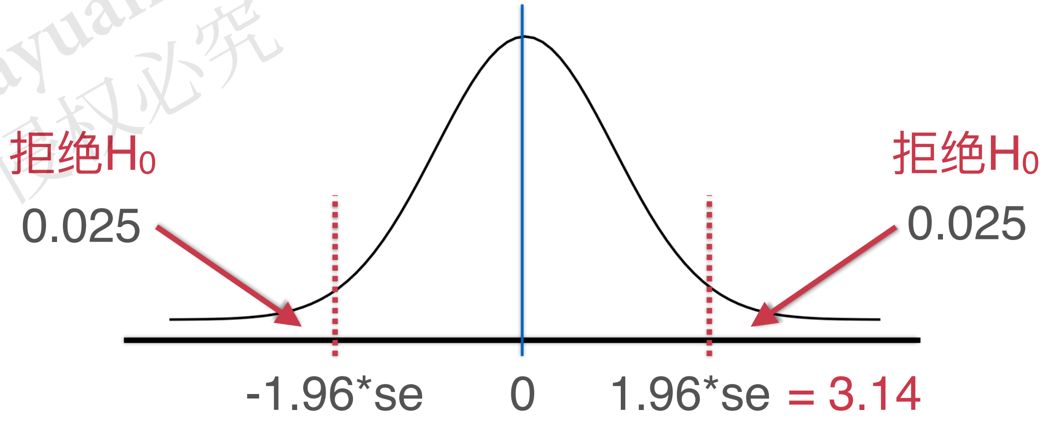
- **如果备择假设是真的，犯二类错误的概率是多少?**
  - H_A为真时，如果真值与零假设参数值非常接近, 则很难检测到差异(即更难拒绝H_0)；如果真值与零假设参数值非常远, 则很容易检测到差异(即更易拒绝H_0)
  - Power = P(拒绝H_0 | H_A 为真) 的大小取决于真值与零假设参数值的距离

统计功效

以往的资料告诉我们, 一班和二班(各50人)数学成绩的标准差均为8; 一班接受数学训练，二班没有接受数学训练，现收集了两班的数学成绩。问两班成绩之差至少为多大可以让我们在显著性水平为0.05时拒绝零假设?(H_0: $\mu_1 - \mu_2 = 0$; H_A: $\mu_1 - \mu_2 \neq 0$)

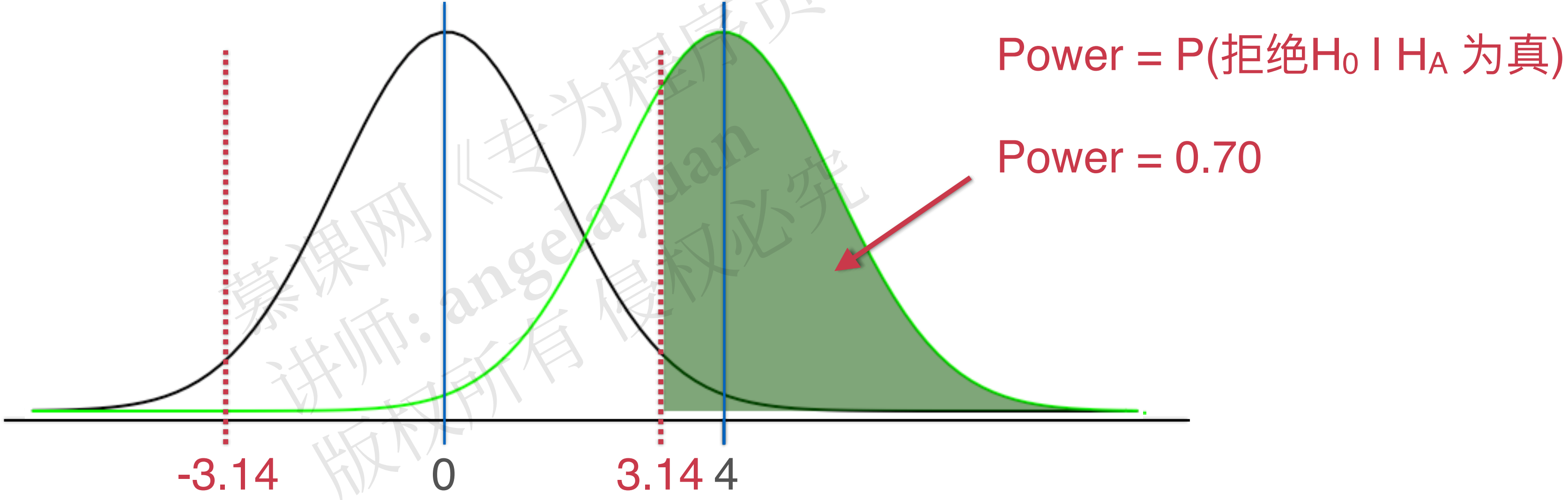
$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2})$$

$$se = \sqrt{\frac{8^2}{50} + \frac{8^2}{50}} = 1.6$$



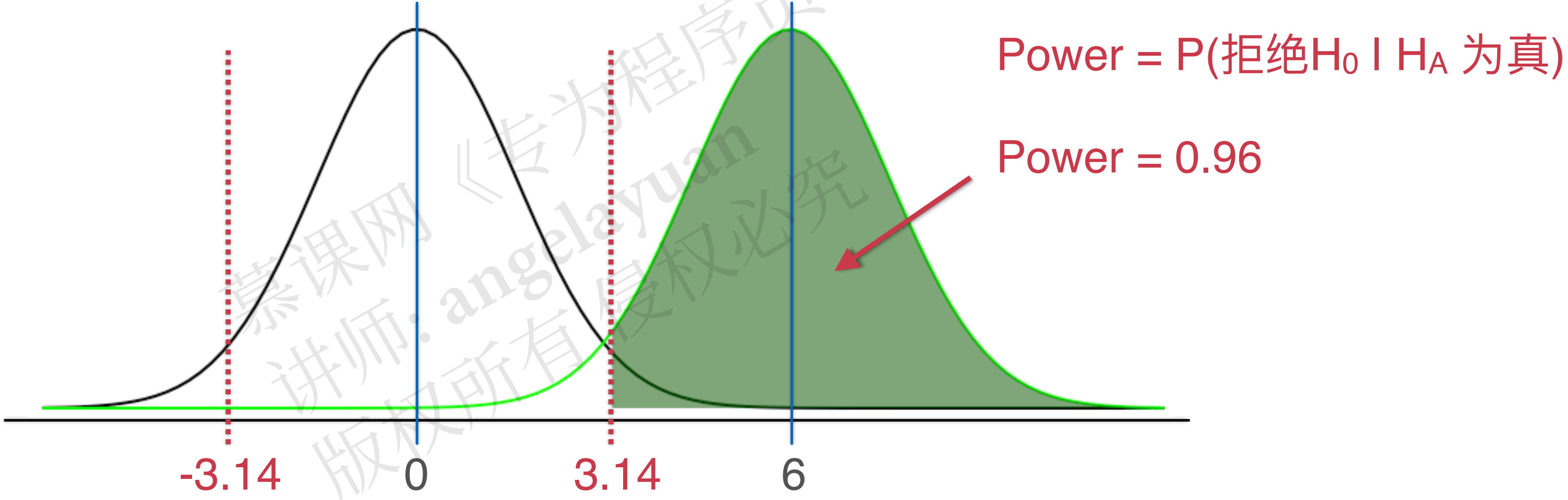
统计功效

如果两班成绩均值差异的真值为4分，基于上一页题干，问能够检测到4分差异的统计功效是多少?



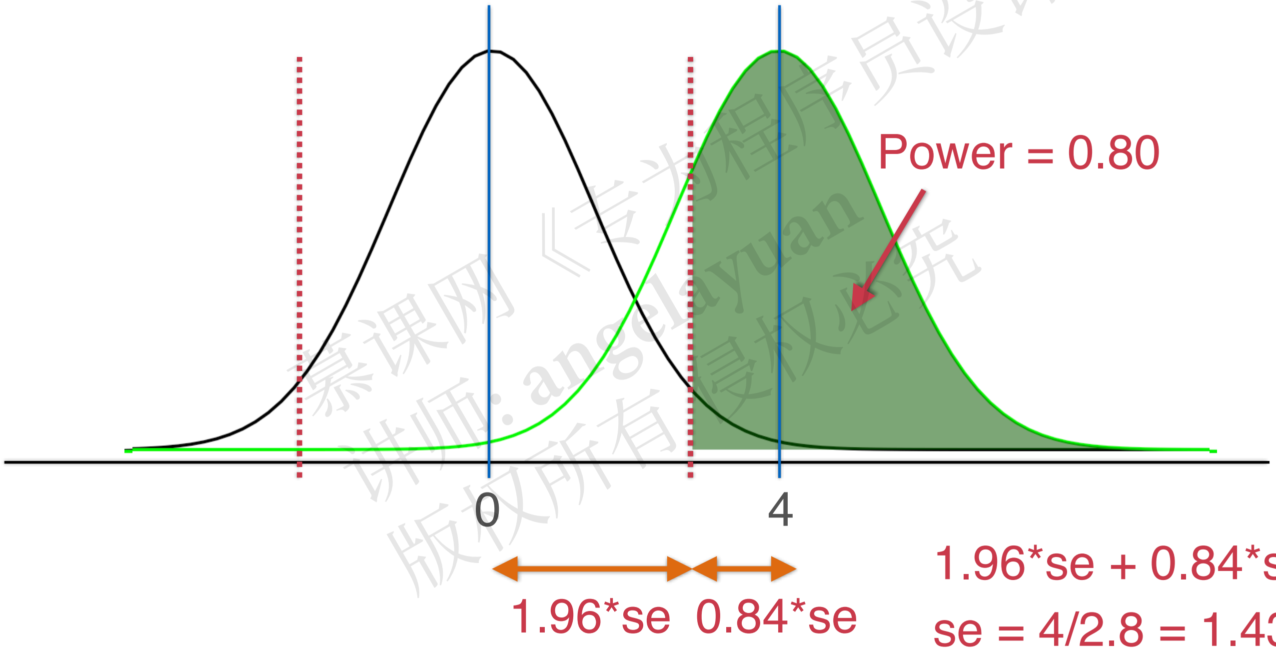
统计功效

如果两班成绩均值差异的真值为6分，基于上一页题干，问能够检测到6分差异的统计功效是多少？



样本容量与统计功效

问样本容量多大可以使得能够检测到4分差异的统计功效是80%



$$se = \sqrt{\frac{8^2}{n} + \frac{8^2}{n}} = 1.43$$

n = 63

1.96*se + 0.84*se = 2.8*se = 4

se = 4/2.8 = 1.43

统计功效，样本容量，α，真值与零值距离

- 当α和样本容量固定时，真值与零假设参数值的间距越大，统计功效越大

查阅资料获取

- 当α和统计功效固定时，真值与零假设参数值的间距越大，所需样本容量越小

统计显著性与实际显著性

统计显著性与实际显著性

一班数学成绩的平均分为90, 二班数学成绩的平均分为89.5, 两班数学成绩分布(总体)的标准差为8, 单尾检验

n = 20 → $$Z = \frac{90 - 89.5}{\sqrt{\frac{8^2}{20} + \frac{8^2}{20}}} = 0.198$$ → p = 0.42 > 0.05

n = 2000 → $$Z = \frac{90 - 89.5}{\sqrt{\frac{8^2}{2000} + \frac{8^2}{2000}}} = 1.976$$ → p = 0.024 < 0.05

点估计与零值的真实差异在大样本时更容易检测到(统计显著)

效应量

- 效应量是量化效应(现象)强度/大小的指标
- 效应量有多种测量方式，效应量的绝对值越大表示效应越强，现象越明显
- Cohen’s D: 一种常用的效应量测量，用于量化两个均值差的效应的大小

效应量

- Cohen’s D: 一种常用的效应量测量，用于量化两个均值差的效应的大小

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S}$$

$$S = \sqrt{\frac{(n_1-1){S_1}^2 + (n_2-1){S_2}^2}{n_1+n_2-2}}$$

d的绝对值	效应大小
0.2	较小
0.5	中等
0.8	较大

统计显著性与实际显著性

一班数学成绩的平均分为90, 二班数学成绩的平均分为89.5, 两班数学成绩分布(总体)的标准差为8

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s} = \frac{90 - 89.5}{8} = 0.063 < 0.2$$

d的绝对值	效应大小
0.2	较小
0.5	中等
0.8	较大

统计显著(statistical significance) ≠ 实际显著(practical significance)

z检验，t检验，卡方检验，F检验的前提条件

总体均值

- 正态总体
  - 组内独立性: 样本是通过随机抽样采集的, 是随机分配到两个组的
  - 组间独立性: 两组互相独立(非配对)
  - 如果是无放回抽样, 则样本容量不能超过总体容量的10%

总体均值

- **非正态总体**
  - 组内和组间独立性
  - 如果是无放回抽样，则样本容量不能超过总体容量的10%
  - 样本容量 ≥ 30; 总体的分布越偏，需要的样本容量越大

总体方差

- **正态总体**
  - 同总体均值检验的前提条件
- **非正态总体**
  - 即使满足所有总体均值检验的前提条件，也不能使用卡方/F检验；需要使用非参数方法

本章小结

- 假设检验
  - 什么是假设检验
  - 频率论 vs 贝叶斯
  - 正态总体均值的假设检验
    - z检验
    - t检验
  - 正态总体方差的假设检验
    - 卡方检验
    - F检验
  - 决策错误, 统计功效, 统计显著性与实际显著性, 各种检验的前提条件