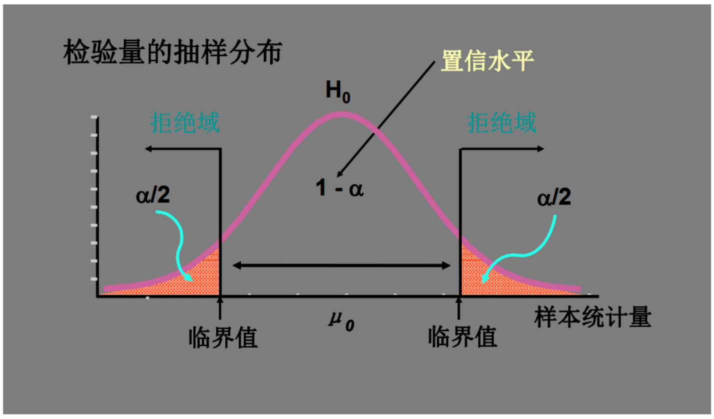
统计检验总结

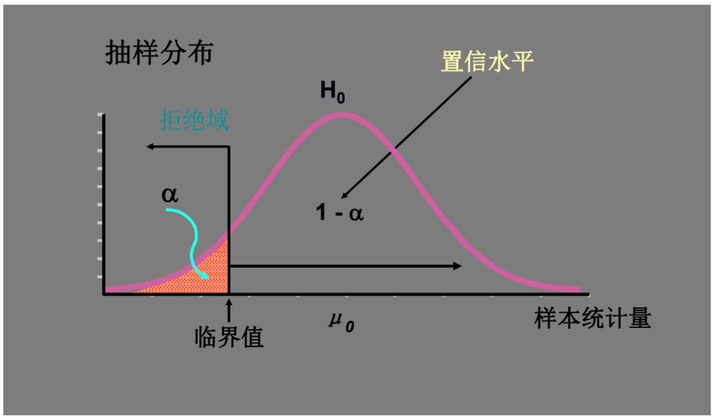
- 由H_1的不同表达形式

假设	研究的问题		
	双侧检验	左侧检验	右侧检验
H_0	$\mu = \mu_0$	$\mu \geq \mu_0$	$\mu \leq \mu_0$
H_1	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$

双侧检验图示(显著性水平与拒绝域)



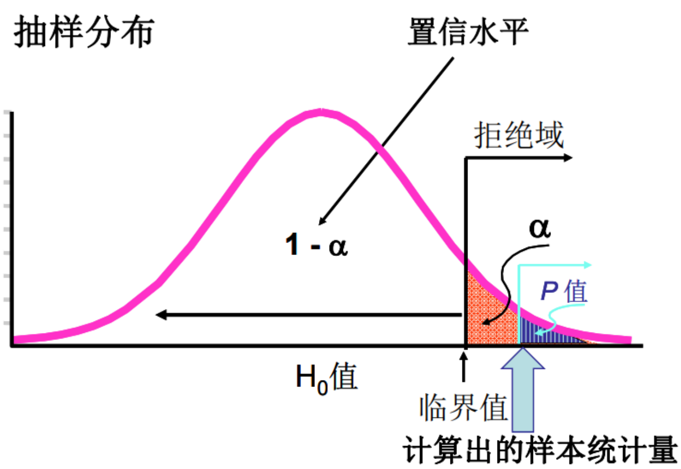
左侧检验图示(显著性水平与拒绝域)



H_0: $\mu \leqslant \mu_0$

H_1: $\mu > \mu_0$

右侧检验的*P* 值



1 Z 检验

1.1 总体方差已知的单总体均值 Z 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$，$\mu \in \mathbb{R}$ 并且 $\sigma^2 > 0$
2. 总体方差σ^2**已知**，但总体期望μ未知
3. 变量$\bar{X}_n$的意思是“独立地从上述分布抽取n个样本后，这个样本集的均值”
4. 从上述$X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$中观测到 n 个样本（无所谓样本大小），记为 $x_1, x_2, \cdots, x_n$
5. 从观测到的样本集$x_1, x_2, \cdots, x_n$可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

$H_0: \mu = \mu_0$ VS. $H_a: \mu \neq \mu_0$

(1) 根据上述前提假设和中心极限定理，可证统计量$\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ 服从$N(0,1)$ 分布。

(2) 现在我们假设 $H_0: \mu = \mu_0$ 是正确的，即总体期望真的为 μ_0. 可知 $\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$ （我们叫做Z 统计量）服从$N(0,1)$ 分布。

(3) 结合之前得到的观测样本，Z 统计量的观测值变成了$\dfrac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$, 可算出一个实数。这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。 p-value 的计算式为：$p - value = P\left(|Z| > \dfrac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right)$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0: \mu = \mu_0$ 是正确的”很可能有问题，我们应该拒绝$H_0: \mu = \mu_0$。

- $H_0: \mu < \mu_0$ VS. $H_a: \mu \geq \mu_0$
- $H_0: \mu \geq \mu_0$ VS. $H_a: \mu < \mu_0$

1.2 总体方差未知的大样本单总体均值 Z 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$，$\mu \in \mathbb{R}$ 并且 $\sigma^2 > 0$
2. 总体方差σ^2**未知**，总体期望μ未知
3. 变量$\bar{X}_n$的意思是"独立地从上述分布抽取n个样本后，这个样本集的均值"。变量S_n的意思是"独立地从上述分布抽取n个样本后，这个样本集的标准差"
4. 从上述$X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$中观测到 n 个样本（观测到大样本），记为 $x_1, x_2, \cdots, x_n$
5. 从观测到的样本集$x_1, x_2, \cdots, x_n$可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$和样本标准差 $s_n = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n-1}}$

$H_0 : \mu = \mu_0$ VS. $H_a : \mu \neq \mu_0$

(1) 根据上述前提假设和中心极限定理，可证统计量$\dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$ 近似服从$N(0,1)$ 分布。

(2) 现在我们假设 $H_0 : \mu = \mu_0$ 是正确的，即总体期望真的为 μ_0. 可知 $\dfrac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$ （我们叫做Z 统计量）服从$N(0,1)$ 分布。

(3) 结合之前得到的观测样本，Z 统计量的观测值变成了$\dfrac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$, 可算出一个实数。这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算"在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率" 来衡量，这个概率我们叫做 p-value。 p-value 的计算式为：$p-value = P\left(|Z| > \dfrac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} \right)$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设" $H_0 : \mu = \mu_0$ 是正确的"很可能有问题，我们应该拒绝$H_0 : \mu = \mu_0$。

1.3 大样本单总体比例 Z 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} Bin(N,\pi)$, $X_i \in \{0,N\}, \pi \in [0,1]$, $N \in \mathbb{Z}$
2. 总体参数N和π皆未知
3. 变量P_n的意思是“独立地从上述分布抽取n个样本后，这个样本集的成功比例（概率）”
4. 从上述$X_1 \cdots X_n \sim^{iid} Bin(N,\pi)$中观测到 n 个样本（观测到大样本）， 记为 $x_1, x_2, \cdots, x_n$
5. 从观测到的样本集$x_1, x_2, \cdots, x_n$可算出样本比例 $p_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n * N}$

$H_0 : \pi = \pi_0$ VS. $H_a : \pi \neq \pi_0$

(1) 根据上述前提假设，可证统计量$\dfrac{P_n - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$近似服从$N(0,1)$ 分布。

(2) 现在我们假设 $H_0 : \pi = \pi_0$ 是正确的，即总体比例真的为 π_0. 可知 $\dfrac{P_n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ （我们叫做Z 统计量）服从$N(0,1)$ 分布。

(3) 结合之前得到的观测样本，Z 统计量的观测值变成了$\dfrac{p_n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$, 可算出一个实数。这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。 p-value 的计算式为：$p - value = P\left(|Z| > \dfrac{p_n - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}\right)$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0 : \pi = \pi_0$ 是正确的”很可能有问题，我们应该拒绝$H_0 : \pi = \pi_0$。

1.4 两总体方差已知的双总体均值差 Z 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma_X^2)$，$Y_1 \cdots Y_m \sim^{iid} N(\mu_Y, \sigma_Y^2)$并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma^2 > 0$
2. 两个分部的总体方差σ_X^2和σ_Y^2**已知**，但两个总体期望μ_X和μ_Y未知
3. 变量$\bar{X}_n$的意思是“独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取 n 个样本后，这个样本集的均值”。变量$\bar{Y}_m$的意思是“独立地从上述$N(\mu_Y, \sigma_Y^2)$分布抽取 m 个样本后，这个样本集的均值”
4. 从上两个分布中分别观测到 n和m个样本， 记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_m$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ 和 $\bar{y}_m = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$

$H_0 : \mu_X - \mu_Y = \mu_0$ VS. $H_a : \mu_X - \mu_Y \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\bar{X}_n - \bar{Y}_m$ 服从$N(\mu_X - \mu_Y, \dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m})$ 分布。

(2) 现在我们假设 $H_0 : \mu_X - \mu_Y = \mu_0$ 是正确的，可知统计量$\bar{X}_n - \bar{Y}_m$ 服从$N(\mu_0, \dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m})$ 分布，也就是说统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$（我们叫做$Z$ 统计量）服从$N(0,1)$分布。

(3) 结合之前观测到的样本， Z 统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$的观测值变成了$\dfrac{(\bar{x}_n - \bar{y}_m) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$，可算出一个实数。这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：

$p - value = P(|Z| > \dfrac{(\bar{x}_n - \bar{y}_m) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}})$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0 : \mu_X - \mu_Y = \mu_0$ 是正确的”很可能有问题，我们应该拒绝$H_0 : \mu_X - \mu_Y = \mu_0$。

1.5 两总体方差未知的大样本双总体均值差 Z 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma_X^2)$，$Y_1 \cdots Y_m \sim^{iid} N(\mu_Y, \sigma_Y^2)$并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma^2 > 0$
2. 两个分部的总体方差σ_X^2和σ_Y^2**未知**，两个总体期望μ_X和μ_Y**未知**
3. 变量$\bar{X}_n$的意思是"独立地从上述$N(\mu_X, \sigma^2)$分布抽取 n 个样本后，这个样本集的均值"。变量S_X的意思是"独立地从上述$N(\mu_X, \sigma^2)$分布抽取n个样本后，这个样本集的标准差"。变量$\bar{Y}_m$的意思是"独立地从上述$N(\mu_Y, \sigma^2)$分布抽取 m 个样本后，这个样本集的均值"。变量S_Y的意思是"独立地从上述$N(\mu_Y, \sigma^2)$分布抽取m个样本后，这个样本集的标准差"。
4. 从上两个分布中分别观测到 n和m个样本 （观测到的两个都是大样本）， 记为 $x_1, x_2, \cdots, x_n$ 和$y_1, y_2, \cdots, y_m$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$，$\bar{y}_m = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$和样本标准差$s_X = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1}}$，$s_Y = \sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}}$

$H_0 : \mu_X - \mu_Y = \mu_0$ VS. $H_a : \mu_X - \mu_Y \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\bar{X}_n - \bar{Y}_m$ 近似服从$N(\mu_X - \mu_Y, \dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m})$ 分布，即 $\dfrac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$服从$N(0,1)$分布

(2) 现在我们假设 $H_0 : \mu_X - \mu_Y = \mu_0$ 是正确的，可知统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$ （我们叫做Z 统计量）服从$N(0,1)$ 分布。

(3) 结合之前观测到的样本， Z 统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$的观测值变成了$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$，可算出一个实数。这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算"在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率" 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：$p - value = P(|Z| > \dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}})$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设" $H_0 : \mu_X - \mu_Y = \mu_0$ 是正确的"很可能有问题，我们应该拒绝$H_0 : \mu_X - \mu_Y = \mu_0$。

1.6 两总体比例差Z检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} Bin(N_x, \pi_X)$， $X_i \in \{0, N_X\}, \pi_X \in [0,1]$， $N_X \in \mathbb{Z}$。
   $Y_1 \cdots Y_m \sim^{iid} Bin(N_Y, \pi_Y)$， $Y_i \in \{0, N_Y\}, \pi_Y \in [0,1]$， $N_Y \in \mathbb{Z}$。
2. 总体参数N_X，N_Y，π_X，π_Y皆未知
3. 变量P_X的意思是“独立地从上述$X_1 \cdots X_n \sim^{iid} Bin(N_x, \pi_X)$分布抽取$n$个样本后，这个样本集的成功比例（概率）”。变量$P_Y$的意思是“独立地从上述$Y_1 \cdots Y_m \sim^{iid} Bin(N_Y, \pi_Y)$分布抽取 m 个样本后，这个样本集的成功比例（概率）”。
4. 从上述$X_1 \cdots X_n \sim^{iid} Bin(N_x, \pi_X)$中观测到 n 个样本， 记为 $x_1, x_2, \cdots, x_n$。从上述$Y_1 \cdots Y_m \sim^{iid} Bin(N_Y, \pi_Y)$中观测到 m 个样本， 记为 $y_1, y_2, \cdots, y_n$。
5. 从观测到的样本集$x_1, x_2, \cdots, x_n$可算出样本比例 $p_X = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$。从观测到的样本集$y_1, y_2, \cdots, y_n$可算出样本比例 $p_Y = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$。

$H_0 : \pi_X - \pi_Y = 0$ VS. $H_0 : p_X - p_Y \neq 0$

(1) 根据上述前提假设，可证统计量$\dfrac{(P_X - P_Y) - (\pi_X - \pi_Y)}{\sqrt{\dfrac{P_X(1-P_X)}{n} + \dfrac{P_Y(1-P_Y)}{m}}}$ 近似服从$N(0,1)$ 分布。

(2) 现在我们假设 $H_0 : \pi_X - \pi_Y = 0$ 是正确的。 可知 $\dfrac{P_X - P_Y}{\sqrt{\dfrac{P_X(1-P_X)}{n} + \dfrac{P_Y(1-P_Y)}{m}}}$ （我们叫做Z 统计量）服从$N(0,1)$ 分布。

(3) 结合之前得到的观测样本，Z 统计量的观测值变成了$\dfrac{p_X - p_Y}{\sqrt{\dfrac{p_X(1-p_X)}{n} + \dfrac{p_Y(1-p_Y)}{m}}}$, 可算出一个实数。

这时候我们需要看一下这个实数在$N(0,1)$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在$N(0,1)$ 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。 p-value 的计算式为：

$$p-value = P\left(|Z| > \frac{p_X - p_Y}{\sqrt{\dfrac{p_X(1-p_X)}{n} + \dfrac{p_Y(1-p_Y)}{m}}} \right)$$

。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“$H_0 : \pi_X - \pi_Y = 0$ 是正确的”很可能有问题，我们应该拒绝$H_0 : \pi_X - \pi_Y = 0$。

2 T检验

2.1 总体方差未知的小样本单总体均值 T 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$，$\mu \in \mathbb{R}$ 并且 $\sigma^2 > 0$
2. 总体方差σ^2和总体期望μ**全部未知**
3. 变量$\bar{X}_n$的意思是“独立地从上述分布抽取n个样本后，这个样本集的均值”。变量S_n的意思是“独立地从上述分布抽取n个样本后，这个样本集的标准差”
4. 从上述$X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$中观测到 n 个样本， 记为 $x_1, x_2, \cdots, x_n$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ 和样本标准差 $s_n = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n-1}}$

$H_0 : \mu = \mu_0$ VS. $H_a : \mu \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$服从自由度为$n-1$的$t$分布。

(2) 现在我们假设 $H_0 : \mu = \mu_0$ 是正确的，即总体期望真的为 μ_0. 可知统计量 $\dfrac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$ （我们叫做 T 统计量）服从 t_{n-1} 分布。

(3) 结合之前得到的观测样本，T 统计量的观测值变成了$\dfrac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$, 可算出一个实数。这时候我们需要看一下这个实数在t_{n-1} 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在t_{n-1} 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：$p-value = P\left(|T| > \dfrac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}} \right)$。如果p-value很小，即这个实数非常极端，说明我们这个Z 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0 : \mu = \mu_0$ 是正确的”很可能有问题，我们应该拒绝$H_0 : \mu = \mu_0$。

2.2 两个样本数相等，且两个总体方差相等的，独立双小样本，双总体均值差 T 检验

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma^2)$，$Y_1 \cdots Y_n \sim^{iid} N(\mu_Y, \sigma^2)$，并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma^2 > 0$
2. 两个分布的总体方差已知相等，记为σ^2，但方差的值未知。两个总体期望可能相等也可能不相等，记为μ_X和μ_Y，参数值也未知。
3. 变量$\bar{X}_n$的意思是"独立地从上述$N(\mu_X, \sigma^2)$分布抽取 n 个样本后，这个样本集的均值"。变量S_X的意思是"独立地从上述$N(\mu_X, \sigma^2)$分布抽取n个样本后，这个样本集的标准差"。变量$\bar{Y}_n$的意思是"独立地从上述$N(\mu_Y, \sigma^2)$分布抽取 n 个样本后，这个样本集的均值"。变量S_Y的意思是"独立地从上述$N(\mu_Y, \sigma^2)$分布抽取n个样本后，这个样本集的标准差"。
4. 从上两个分布中相互独立地分别观测到 n和n个样本，记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_n$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ 和 $\bar{y}_n = \dfrac{y_1 + y_2 + \cdots + y_n}{n}$。样本标准差$s_X = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1}}$和$s_Y = \sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}}$

$H_0: \mu_X - \mu_Y = \mu_0$ VS. $H_a: \mu_X - \mu_Y \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\dfrac{(\bar{X}_n - \bar{Y}_n) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2)}{(n+n-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{n}}}$服从自由度为$2n-2$的 t 分布。

(2) 现在我们假设 $H_0: \mu_X - \mu_Y = \mu_0$ 是正确的，可知统计量$\dfrac{(\bar{X}_n - \bar{Y}_n) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2)}{(n+n-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{n}}}$（我们叫做 T 统计量）服从自由度为$2n-2$的 t 分布。

(3) 结合之前观测到的样本，统计量$\dfrac{(\bar{X}_n - \bar{Y}_n) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2)}{(n+n-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{n}}}$的观测值变成了$\dfrac{(\bar{x}_n - \bar{y}_n) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{n}(y_i - \bar{y}_n)^2)}{(n+n-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{n}}}$，可算出一个实数。这时候我们需要看一下这个实数在$t_{n+n-2}$分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算"在t_{n+n-2}分布中观测到一个比这个实数更加极端的值的概率" 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：$p - value = P\left(|T| > \dfrac{\bar{x}_n - \bar{y}_n - \mu_0}{\sqrt{\dfrac{2}{n} \cdot \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n+n-2}}}\right)$。如果p-value很小，即这个实数非常极端，说明我们这个T 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设" $H_0: \mu_X - \mu_Y = \mu_0$ 是正确的"很可能有问题，我们应该拒绝$H_0: \mu_X - \mu_Y = \mu_0$。

2.3 两个样本数不相等，但两个总体方差相等的情况下**独立双样本 T 检验**

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma^2)$，$Y_1 \cdots Y_m \sim^{iid} N(\mu_Y, \sigma^2)$，并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma^2 > 0$
2. 两个分布的总体方差已知相等，记为σ^2，但方差的值未知。两个总体期望可能相等也可能不相等，记为μ_X和μ_Y，两个期望值也未知。
3. 变量$\bar{X}_n$的意思是“独立地从上述$N(\mu_X, \sigma^2)$分布抽取 n 个样本后，这个样本集的均值”。变量S_X的意思是“独立地从上述$N(\mu_X, \sigma^2)$分布抽取n个样本后，这个样本集的标准差”。变量$\bar{Y}_m$的意思是“独立地从上述$N(\mu_Y, \sigma^2)$分布抽取 m 个样本后，这个样本集的均值”。变量S_Y的意思是“独立地从上述$N(\mu_Y, \sigma^2)$分布抽取m个样本后，这个样本集的标准差”。
4. 从上两个分布中相互独立地分别观测到 n和m个样本， 记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_m$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ 和 $\bar{y}_m = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$。样本标准差$s_X = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1}}$和$s_Y = \sqrt{\dfrac{\sum_{i=1}^{m}(y_i - \bar{y}_m)^2}{m-1}}$

$H_0: \mu_X - \mu_Y = \mu_0$ VS. $H_a: \mu_X - \mu_Y \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2)}{(n+m-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$服从自由度为 $n + m - 2$的 t 分布。

(2) 现在我们假设 $H_0: \mu_X - \mu_Y = \mu_0$ 是正确的，可知统计量 $\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2)}{(n+m-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$（我们叫做 T 统计量）服从自由度为 $n + m - 2$的 t 分布。

(3) 结合之前观测到的样本，统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2)}{(n+m-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$的观测值变成了 $\dfrac{(\bar{x}_n - \bar{y}_m) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{m}(y_i - \bar{y}_m)^2)}{(n+m-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}}$，可算出一个实数。这时候我们需要看一下这个实数在$t_{n+m-2}$分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在 t_{n+m-2} 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：$p - value = P\left(|T| > \dfrac{(\bar{x}_n - \bar{y}_m) - \mu_0}{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + \sum_{i=1}^{m}(y_i - \bar{y}_m)^2)}{(n+m-2)}} \cdot \sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \right)$。如果p-value很小，即这个实数非常极端，说明我们这个T 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0: \mu_X - \mu_Y = \mu_0$ 是正确的”很可能有问题，我们应该拒绝 $H_0: \mu_X - \mu_Y = \mu_0$。

2.4 两个样本数不相等，且两个总体方差不相等的情况下**独立双样本 T 检验 （即Welch检验）**

前提假设：

1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma_X^2)$，$Y_1 \cdots Y_m \sim^{iid} N(\mu_Y, \sigma_Y^2)$并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma_X^2 > 0$，$\sigma_Y^2 > 0$
2. 两个分布的总体方差σ_X^2和σ_Y^2和两个总体期望μ_X和μ_Y**全部未知**
3. 变量$\bar{X}_n$的意思是“独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取 n 个样本后，这个样本集的均值”。变量S_X的意思是“独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取n个样本后，这个样本集的标准差”。变量$\bar{Y}_m$的意思是“独立地从上述$N(\mu_Y, \sigma_Y^2)$分布抽取 m 个样本后，这个样本集的均值”。变量S_Y的意思是“独立地从上述$N(\mu_Y, \sigma_Y^2)$分布抽取m个样本后，这个样本集的标准差”。
4. 从以上两个分布中相互独立地分别观测到 n和m个样本， 记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_m$
5. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ 和 $\bar{y}_m = \dfrac{y_1 + y_2 + \cdots + y_m}{m}$。样本标准差$s_X = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n-1}}$和$s_Y = \sqrt{\dfrac{\sum_{i=1}^{m} (y_i - \bar{y}_m)^2}{m-1}}$

$H_0 : \mu_X - \mu_Y = \mu_0$ VS. $H_a : \mu_X - \mu_Y \neq \mu_0$

(1) 根据上述前提假设，可证统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - (\mu_Y - \mu_X)}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$服从自由度为 $df = \dfrac{\left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)^2}{\dfrac{\left(\frac{S_X^2}{n}\right)^2}{n-1} + \dfrac{\left(\frac{S_Y^2}{m}\right)^2}{m-1}}$ 的 t 分布。

(2) 现在我们假设 $H_0 : \mu_X - \mu_Y = \mu_0$ 是正确的，可知统计量$\dfrac{(\bar{X}_n - \bar{Y}_m) - \mu_0}{\sqrt{\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}}}$服从自由度为的 t

$$df = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{\left(\frac{S_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_Y^2}{m}\right)^2}{m-1}}$$

分布。.....

(3) ……

2.5 配对双样本 T 检验

前提假设：

1. 两个样本集里的数据是一一配对的，或者是同一个物体前后两次测量结果。两个样本大小肯定一样。
2. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma_X^2)$，$Y_1 \cdots Y_n \sim^{iid} N(\mu_Y, \sigma_Y^2)$并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma_X^2 > 0$，$\sigma_Y^2 > 0$
3. 两个分布的总体方差σ_X^2和σ_Y^2和两个总体期望μ_X和μ_Y**全部未知**
4. 变量$\bar{D}_n$的意思是“独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取 n 个样本，并从上述$N(\mu_Y, \sigma_Y^2)$分布抽取这 n 个样本对应的版本后，这个两个样本集的差的均值”。变量S_D的意思是“独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取 n 个样本，并从上述$N(\mu_Y, \sigma_Y^2)$分布抽取 这 n 个样本对应的版本后，这个两个样本集的差的标准差”。参数D的意思是上述两个总体一一对应的差值的总体期望。
5. 从上两个分布中观测到 n和n个样本， 记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_n$。并且可算出两个样本之间一一对应的差值$d_i = x_i - y_i$，记为$d_1, d_2, \cdots, d_n$
6. 从观测到的样本集可算出样本平均数 $\bar{x}_n = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$，$\bar{y}_n = \dfrac{y_1 + y_2 + \cdots + y_n}{n}$。样本标准差$s_X = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1}}$，$s_Y = \sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}}$。样本差值均值 $\bar{d}_n = \dfrac{d_1 + d_2 + \cdots + d_n}{n}$。样本差值标准差$s_D = \sqrt{\dfrac{\sum_{i=1}^{n}(d_i - \bar{d}_n)^2}{n-1}}$。

$H_0 : D = D_0$ VS. $H_0 : D \neq D_0$

(1) 根据上述前提假设，可证统计量$\dfrac{\bar{D}_n - D}{S_D/\sqrt{n}}$服从自由度为 $df = n - 1$ 的 t 分布。

(2) 现在我们假设 $H_0 : D = D_0$ 是正确的，可知统计量$\dfrac{\bar{D}_n - D_0}{S_D/\sqrt{n}}$（我们叫做 T 统计量）服从自由度为 $n - 1$的 t 分布。

(3) 结合之前观测到的样本，统计量$\dfrac{\bar{D}_n - D_0}{S_D/\sqrt{n}}$的观测值变成了$\dfrac{\bar{d}_n - D_0}{s_D/\sqrt{n}}$，可算出一个实数。这时候我们需要看一下这个实数在$t_{n-1}$分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在t_{n-1} 分布中观测到一个比这个实数更加极端的值的概率” 来衡量，这个概率我们叫做 p-value。p-value 的计算式为：$p - value = P\left(|T| > \dfrac{\bar{d}_n - D_0}{s_D/\sqrt{n}}\right)$。如果p-value很小，即这个实数非常极端，说明我们这个T 统计量的观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0 : D = D_0$ 是正确的”很可能有问题，我们应该拒绝$H_0 : D = D_0$。

3 χ^2检验

3.1 单总体方差 χ^2 检验

前提假设：

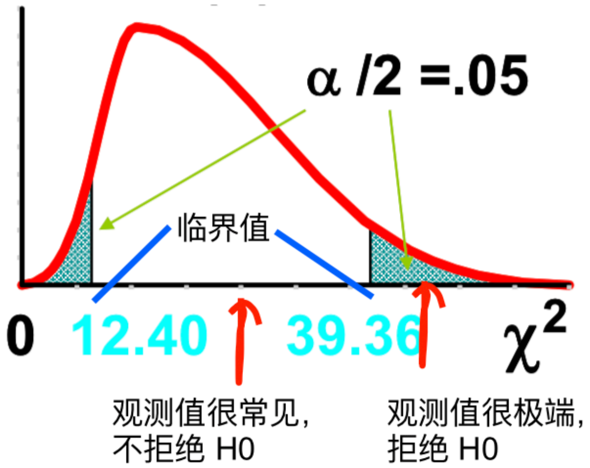
1. $X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$，$\mu \in \mathbb{R}$ 并且 $\sigma^2 > 0$
2. 总体方差σ^2和总体期望μ**全部未知**
3. 变量$\bar{X}_n$的意思是“独立地从上述分布抽取n个样本后，这个样本集的均值”。变量S_n的意思是“独立地从上述分布抽取n个样本后，这个样本集的标准差”
4. 从上述$X_1 \cdots X_n \sim^{iid} N(\mu, \sigma^2)$中观测到 n 个样本，记为 $x_1, x_2, \cdots, x_n$
5. 从观测到的样本集可算出样本标准差$s_n = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n-1}}$

$H_0: \sigma = \sigma_0$ VS. $H_a: \sigma \neq \sigma_0$

(1) 根据上述前提假设，可证统计量$\dfrac{(n-1)S_n^2}{\sigma^2}$服从自由度为$n-1$的卡方分布。

(2) 现在我们假设 $H_0: \sigma = \sigma_0$ 是正确的，即总体标准差真的为 σ_0. 可知统计量 $\dfrac{(n-1)S_n^2}{\sigma_0^2}$ （我们叫做 χ^2统计量）服从 χ^2_{n-1} 分布。

(3) 结合之前得到的观测样本，χ^2统计量的观测值变成了$\dfrac{(n-1)s_n^2}{\sigma_0^2}$, 可算出一个实数。这时候我们需要看一下这个实数在χ^2_{n-1} 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算“在χ^2_{n-1} 分布中，哪两个临界值之外的区域各为$\dfrac{\alpha}{2}$”来判断。如果χ^2统计量的观测值落在临界



值之外，说明这个观测值其实是一个很少见的极端值，那么我们最开始做出的假设“ $H_0: \sigma = \sigma_0$ 是正确的”很可能有问题，我们应该拒绝$H_0: \sigma = \sigma_0$。

4 F检验

4.1 两总体方差比 F检验

前提假设：

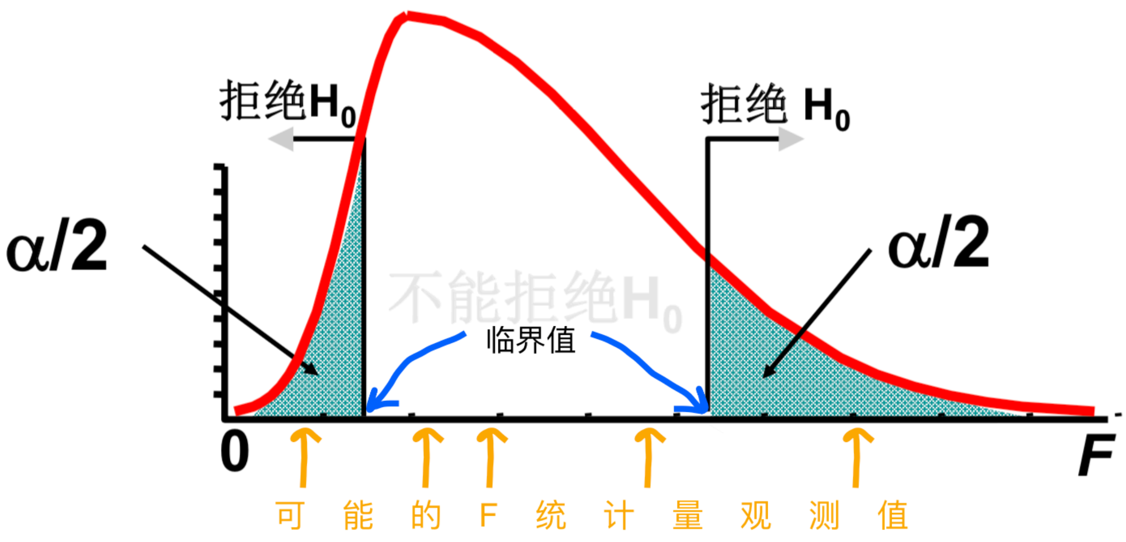
1. $X_1 \cdots X_n \sim^{iid} N(\mu_X, \sigma_X^2)$，$Y_1 \cdots Y_m \sim^{iid} N(\mu_Y, \sigma_Y^2)$并且 $\mu_X \in \mathbb{R}$，$\mu_Y \in \mathbb{R}$，$\sigma_X^2 > 0$，$\sigma_Y^2 > 0$
2. 两个分布的总体方差σ_X^2和σ_Y^2和两个总体期望μ_X和μ_Y**全部未知**
3. 变量$\bar{X}_n$的意思是"独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取 n 个样本后，这个样本集的均值"。变量S_X的意思是"独立地从上述$N(\mu_X, \sigma_X^2)$分布抽取n个样本后，这个样本集的标准差"。变量$\bar{Y}_m$的意思是"独立地从上述$N(\mu_Y, \sigma_Y^2)$分布抽取 m 个样本后，这个样本集的均值"。变量S_Y的意思是"独立地从上述$N(\mu_Y, \sigma_Y^2)$分布抽取m个样本后，这个样本集的标准差"。
4. 从以上两个分布中相互独立地分别观测到 n和m个样本，记为 $x_1, x_2, \cdots, x_n$和$y_1, y_2, \cdots, y_m$。
5. 从观测到的样本集可算出样本标准差$s_X = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{n-1}}$和$s_Y = \sqrt{\frac{\sum_{i=1}^m (y_i - \bar{y}_m)^2}{m-1}}$

$H_0 : \sigma_X = \sigma_Y$ VS. $H_a : \sigma_X \neq \sigma_Y$

(1) 根据上述前提假设，可证统计量$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$服从自由度为$n-1$和$m-1$的 F分布。

(2) 现在我们假设 $H_0 : \sigma_X = \sigma_Y$ 是正确的，两个即总体标准差真的相同. 可知统计量 $\frac{S_X^2}{S_Y^2}$ （我们叫做 F 统计量）服从 $F_{n-1,m-1}$ 分布。

(3) 结合之前得到的观测样本，F统计量的观测值变成了$\frac{s_X^2}{s_Y^2}$, 可算出一个实数。这时候我们需要看一下这个实数在$F_{n-1,m-1}$ 分布的什么位置？非常常见还是非常极端？这个实数是否极端这件事可以通过计算"在$F_{n-1,m-1}$ 分布中，哪两个临界值之外的区域各为$\frac{\alpha}{2}$"来判断。如果F统计量的观测值落在临界



值之外，说明这个观测值其实是一个很少见的极端值，那么我们最开始做出的假设" $H_0 : \sigma_X = \sigma_Y$ 是正确的"很可能有问题，我们应该拒绝$H_0 : \sigma_X = \sigma_Y$。