3.2　抽样分布

3.2 抽样分布

❖研究总体与从中抽取的样本之间的关系是统计学的中心内容。

❖对这种关系的研究可从两方面着手：

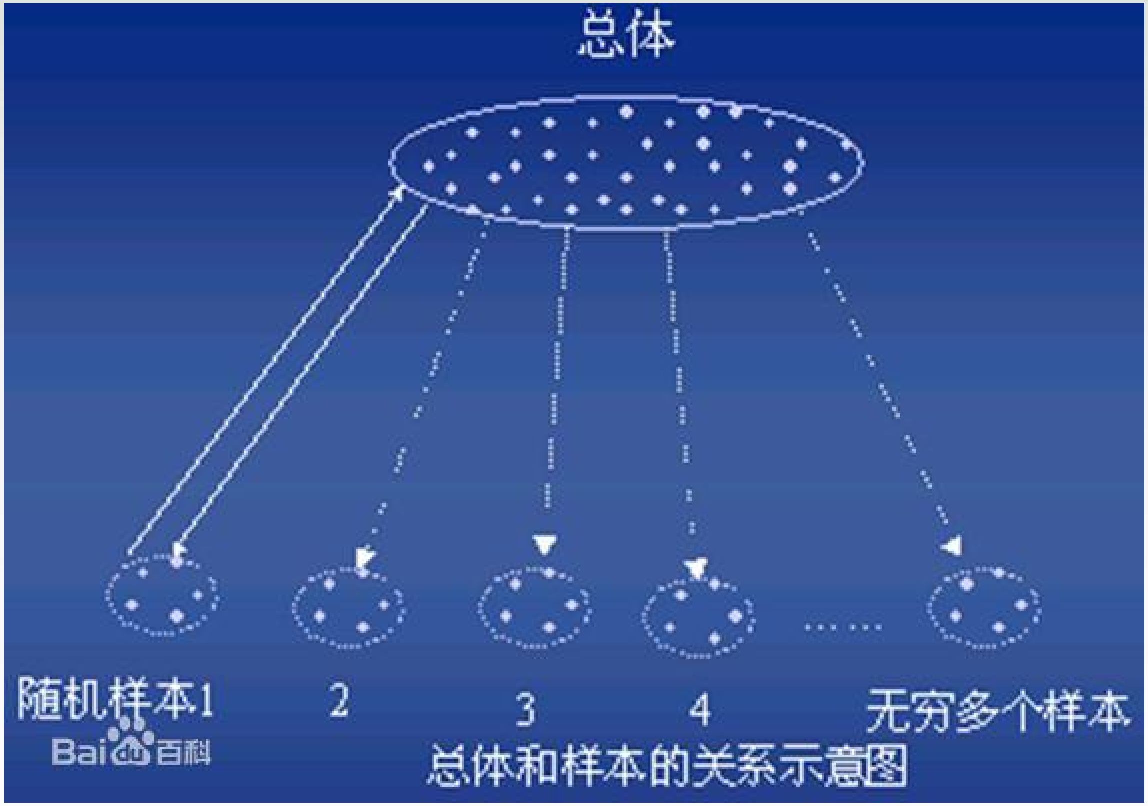
- 从总体到样本，也即**抽样分布**（sampling distribution）；
- 从样本到总体，也即**统计推断**（statistical inference）。

❖为正确利用样本去推断总体，并能正确地理解统计推断的结论，须对样本的抽样分布有所了解。

❖由总体中随机地抽取若干个体组成样本，即使每次抽取的样本含量相同，其统计量（$\bar{x}, S$）也将随样本的不同而有所不同，因而样本统计量也是随机变量，也有其概率分布。

❖我们把统计量的概率分布称为**抽样分布**。（视频）

我们全班同学的平均身高是多少？



- 设有一个总体，总体均数为μ，方差为σ^2，总体中各变数为x，将此总体称为原总体。
- 现从这个总体中随机抽取含量为n的样本，样本平均数记为$\bar{x}$。可以设想，从原总体可抽出很多，甚至无穷多个含量为n的样本。
- 由这些样本算得的平均数有大有小，不尽相同，与原总体均数μ相比往往表现出不同程度的差异。
- 这种差异是由随机抽样造成的，称为抽样误差（sampling error）。
- 显然，样本平均数也是一个随机变量，**其概率分布称为样本平均数的抽样分布**。

❖由样本平均数 $\bar{x}$ 构成的总体称为样本平均数的抽样总体，其平均数和标准差分别记为 $\mu_{\bar{x}}$和$\sigma_{\bar{x}}$ 。

❖其中，$\sigma_{\bar{x}}$ 是样本平均数抽样总体的标准差，简称标准误差（standard error），它表示平均数抽样误差的大小。

❖统计学上 $\bar{x}$ 总体的两个参数与x总体的两个参数有什么关系？（讨论）

$$\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

X变量与 $\bar{x}$ 变量分布间关系：

- 若随机变量x服从正态分布N（μ,σ²），x_1，$x_2 \ldots x_n$是由总体的来的随机样本，则统计量 $\bar{x} = \sum x/n$ 的概率分布也是正态分布，且有 $\mu_{\bar{x}} = \mu, \sigma_{\bar{x}}^2 = \sigma^2/n$ 即 $\bar{x}$ 服从正态分布N（μ,σ²）。

- 若随机变量x服从平均数是μ和方差是σ²的分布（不是正态分布），x_1，$x_2 \ldots x_n$是由总体的来的随机样本，则统计量 $\bar{x}$ 的概率分布，在当n相当大时，趋于正态分布。也即**中心极限定理**。

❖上述两个定理说明，样本平均数的抽样分布服从或逼近正态分布。

❖中心极限定理明确告诉我们，不论x服从何种分布，一般只要n>30，就可以认为其平均数的分布是正态的。若x的分布不很偏斜，在n>20时，平均数的分布就近似于正态分布了。

❖这就是正态分布较之其他分布应用更为广泛的原因。

3.2.2 均数标准误

- **均数标准误**（平均数抽样总体的标准差）$\sigma_x = \sigma\sqrt{n}$的大小反映样本数平均数的抽样误差的大小，即精确性的高低。
- 标准差大，说明各样本数的平均数之间的差异程度大，样本平均数的精确性低；反之，其值小，说明平均数之间的差异程度小，样本平均数的精确性高。
- 标准差的大小与原总体的标准差成正比，与样本含量n的平方根成反比。
- **从某一特定总体抽样，因为其标准差是一常数，所以只有增大样本含量才能降低样本均数的平均数的抽样误差。**

❖在实际工作中，总体标准差σ往往是未知的，因而无法求得σ_x。

❖此时，可用样本标准差S估计σ。于是，以S/√n估计σ_x。记S/√n为S_x，称为**样本均数标准误**或者**均数标准误**。

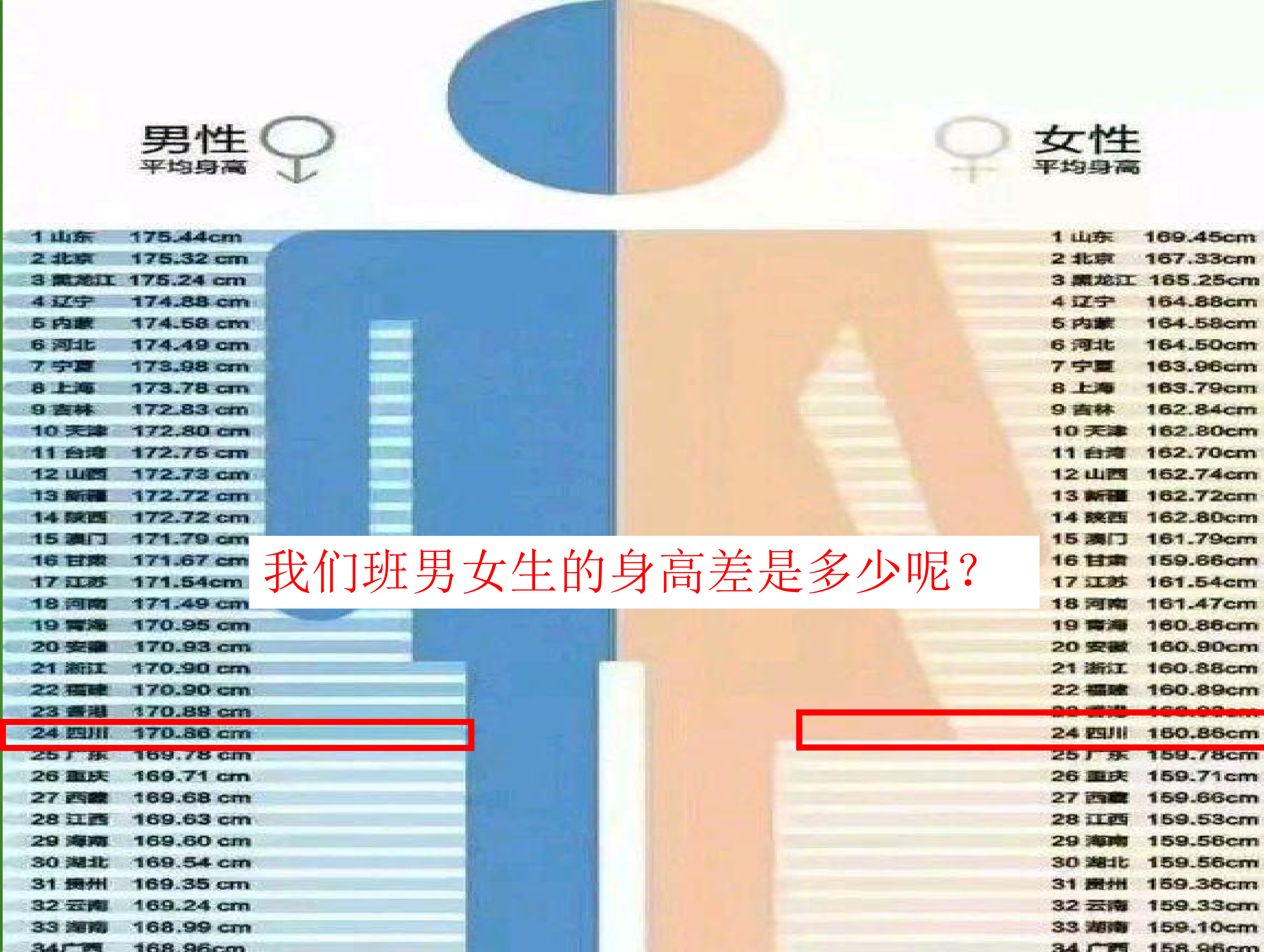
❖S_x是平均数抽样误差的估计值。若样本中各观测值为x_1，x_2，…，x_n，则：

$$S_{\bar{x}}=\frac{S}{\sqrt{n}}=\sqrt{\frac{\sum(x-\bar{x})^2}{n(n-1)}}=\sqrt{\frac{\sum x^2-(\sum x)^2/n}{n(n-1)}}$$

- 注意，样本标准差与样本均数标准误是既有联系又有区别的两个统计量，上式已表明了两者的联系。

- 两者区别在于：样本标准差S是反映样本中各变数x_1，x_2，…，x_n变异程度大小的一个指标。它的大小说明了$\bar{x}$对该样本代表性的强弱。

- 样本均数标准误$S_{\bar{x}}$是样本平均数$\bar{x}_1$，$\bar{x}_2$，…，$\bar{x}_n$的标准差。它是$\bar{x}$抽样误差的估计值，其大小说明了样本间变异程度的大小及$\bar{x}$精确性的高低。



我们班男女生的身高差是多少呢？

3.2.3 两样本均数差数的抽样分布

❖关于两样本均数差数的抽样分布有以下规律：

❖设 $x_1 \sim N(\mu_1, \sigma_1^2)$，$x_2 \sim N(\mu_2, \sigma_2^2)$ 且x_1与x_2相互独立，若从这两个总体里抽取素有可能的样本对（无论样本容量n_1、n_2大小），则样本平均数之差 $\bar{x}_1 - \bar{x}_2$ 服从正态分布，即

$$(\bar{x}_1 - \bar{x}_2) \sim N(\mu_{\bar{x}_1 - \bar{x}_2}, \sigma^2_{\bar{x}_1 - \bar{x}_2})$$

❖总体参数有如下关系： （式1）

$$\begin{cases} \mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 \\ \sigma^2_{\bar{x}_1 - \bar{x}_2} = \sigma_1^2 / n_1 + \sigma_2^2 / n_2 \end{cases}$$

- 若所有样本对来自同一正态总体 $x \sim N(\mu, \sigma^2)$，则其平均数差数的抽样分布服从正态分布，且 （式2）

$$\begin{cases} \mu_{\bar{x}_1\text{-}\bar{x}_2} = 0 \\ \sigma^2_{\bar{x}_1\text{-}\bar{x}_2} = \sigma^2/n_1 + \sigma^2/n_2 \end{cases}$$

- 若所有样本对来自同一非正态的总体，则其平均数差数的抽样分布按中心极限定理在n_1和n_2相当大时（大于30）时，才逐渐接近于正态分布。参数间的关系同式2。

- 若所有样本对来自两个非正态的总体，且σ_1^2和σ_2^2相差很大时，则其平均数差数的抽样分布很难确定；当σ_1^2和σ_2^2相差不大，且n_1和n_2趋于无穷大时均数差数的抽样分布逐渐趋于正态分布，参数间的关系同式1。

3.2.4 样本均数差数标准误

- 实际中，σ_1^2和σ_2^2常是未知的。但在样本含量充分大的情况下，通常用S_1^2和S_2^2分别代替σ_1^2和σ_2^2。
- 于是 $\sigma_{\bar{x}_1-\bar{x}_2}^2$ 常用$\sqrt{S_1^2/n_1+S_2^2/n_2}$估计，记为：

$$S_{\bar{x}_1-\bar{x}_2}=\sqrt{S_1^2/n_1+S_2^2/n_2}$$

- 上式中，S_1^2和S_2^2分别是样本含量为n_1和n_2的两个样本方差。如果它们所估计的各自总体方差σ_1^2和σ_2^2相等。那么S_1^2和S_2^2就都是σ^2的估计值。此时，应将S_1^2和S_2^2的加权平均值S_0^2作为σ^2的估计值较为合理。

第 4 章
统 计 假 设 检 验

- **统计推断** 是指根据样本以及问题的条件和假定模型对未知事物(即总体)作出的**以概率形式表述的**推断，它主要包括**统计假设检验**和**参数估计**两个内容。(视频)
- **统计假设检验**又叫**显著性检验**

第1节　统计假设检验概述

一、显著性检验的意义和基本原理

二、两种假设

三、显著水平与两类错误

四、双侧检验与单侧检验

五、显著性检验的基本步骤

为什么我们要学习显著性检验？

跑步、跳健身操、广场舞、晚上不吃饭、奶茶用代糖……

我们班的女生从本月开始坚持每天跑步，以上个月的体重为对照来进行试验。已知在上个月我们班女生的体重$\mu_0=50$Kg（已知总体均值），并从长期观察获得标准差$\sigma=5$，从这个月开始我们班的女生坚持每天跑步后2个月，抽得25个女生，其体重平均数为$\overline{x}=48$Kg，问：每天坚持跑步是否有助于减肥呢？

我们知道：

坚持跑步后的25个女生的体重平均值 $\bar{x}$

实验要求：

坚持跑步后全班女生总体的体重平均数 μ.

- 目的　总体平均数（$\mu=\mu_0$）
- 对象　样本平均数 $\bar{x}$

样本平均数=总体平均数？

样本平均数和总体平均数存在差异的原因是什么？

（讨论）

（三） 基本原理及思路

观察值由两部分组成即：$x_i = \mu + \varepsilon_i$ 试验误差

若样本含量为n，则可得到n个观察值

$x_1, x_2, \cdots x_n$ 样本平均数 $\bar{x} = \mu + \bar{\varepsilon}$ 。

说明样本均数并非总体均数，它还包含试验误差的成分。

对于接受不同处理的两个样本来说，则有：

$$\bar{x}_1 = \mu + \bar{\varepsilon}_1 \quad \bar{x}_2 = \mu + \bar{\varepsilon}_2$$

$$\underbrace{\bar{x}_1 - \bar{x}_2}_{\text{两样本均数之差为试验表面效应}} = \underbrace{(\mu_1 - \mu_2)}_{\text{试验的处理效应}} + \underbrace{(\bar{\varepsilon}_1 - \bar{\varepsilon}_2)}_{\text{试验误差}}$$

处理效应 $(\mu_1 - \mu_2)$ 未知，但试验表面效应 $(\bar{x}_1 - \bar{x}_2)$ 是可以计算的，借助于统计方法，试验误差 $(\bar{\varepsilon}_1 - \bar{\varepsilon}_2)$ 也是可以估计的．因此可**从试验的表面效应与试验误差的权衡比较中间接地推断处理效应是否存在**，这就是显著性检验的基本思想。

我们的减肥试验中：

1.如果处理效应不存在，$(\mu_1-\mu_2)=0$
则表面效应 $\bar{x}-\mu_0=2$ 是由误差造成的，说明跑步前后的体重没有显著差异。

2.如果处理效应存在，$(\mu_1-\mu_2)\neq 0$
则表面效应 $\bar{x}-\mu_0=2$ 是由处理效应造成的，说明跑步前后的体重有显著差异

- 统计假设检验在科学研究中是一种非常重要的统计分析方法。在科研试验中常会有两个处理的比较试验，如两种工艺方法的比较；一种新添加剂与对照两处理的比较；两种食品内含物测定方法的比较；检验某产品是否达到某项质量标准，或检测某种有害物质含量是否超标等。
- 对于上述试验数据，均可采用统计假设检验来分析，从而保证获得相对正确可靠的结论。

二、统计假设检验的基本原理

（1）首先对试验样本所在的总体作假设。

1.零假设

总体平均数是未知的，为了得到对总体平均数的推断，可以假设总体平均数$\mu=\mu_0$或$\mu-\mu_0=0$，其意义是试验的表面效应系试验误差，处理无效，故称为**无效假设**，也称为**零假设**，记作H_0，$H_0:\mu=\mu_0$或$H_0:\mu-\mu_0=0$

无效假设是被检验的假设，通过检验可能被接受，也可能被否定。

2. 备择假设：

记为H_A，是在无效假设被否定，拒绝H_0的情况下的所有可供选择的假设

若H_0：$\mu \neq \mu_0$，则备择假设包括以下三种：

$$H_A:\ \mu \neq \mu_0$$

$$H_A:\ \mu > \mu_0$$

$$H_A:\ \mu < \mu_0$$

（2）在无效假设H_0成立的前提下，构造合适的统计量，并由该统计量的抽样分布计算样本统计量的概率。

无效假设H_0：$\mu=\mu_0$ 成立，说明 $\bar{X}-\mu_0$之间的差异是由误差造成的。此时可把试验中所获得的$\bar{X}$看成是从μ_0总体中随机抽取的一个样本均数。由样本均数抽样分布理论可知，从一个均数为μ_0方差为σ^2的正态总体中抽样，所得的一系列样本均数 的分布呈正态分布 $N(\mu_0,\frac{\sigma^2}{n})$ 。

对$\overline{X}$作标准化，则有

$$u=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$$

由上式即可计算出样本统计量u值，并估计出H_0条件下｜u｜超过样本实得值的概率。

由正态分布的双侧分位数（u_{α}）表可知:

$$P(|u|\geq 1.96)=0.05$$

$$P(|u|\geq 2.58)=0.01$$

（3）根据估计出的统计量的概率值大小，做出接受或否定无效假设的推断。

如果估计出的统计量的概率值非常小，说明无效假设H_0认为 $\overline{X}$与μ_0 的差异纯属误差造成的为小概率事件。因而原先所作的无效假设H_0是不正确的，应予以否定，转而接受备择假设H_1；反之，如果估计出的统计量的概率值不是很小，说明 $\overline{X}$与μ_0 纯属误差造成的情况有较大可能会出现，此时的无效假设H_0很可能是正确的，不能被否定。

在跑步减肥试验中：

1.提出假设：

H_0： $\mu=\mu_0=50$，即跑步前后女生的平均体重相等。

H_A： $\mu\neq\mu_0=50$，即跑步前后女生的平均体重不相等。

2.在无效假设H_0成立的前提下，构建合适的统计量，并由该统计量的抽样分布计算样本统计量的概率。

无效假设成立，说明 $\bar{x}$ 是已知 μ_0=总体中随机抽出的一个样本平均数。由抽样分布的理论可知，样本平均数的分布服从原总体的正态分布。

$\mu_0=50$， $\bar{x}=48$，$n=25$，$\sigma=5$，代入标准化的公式得到 $\mu=2$， $1.96<|\mu|<2.58$，所以可推知其概率 $P(|\mu|\geq 2)=p$， $0.01<p<0.05$.

本试验的表面效应 $\bar{x}-\mu_0=2$ 完全由误差造成的概率在0.01～0.05之间。

我们需要注意：

（一）小概率原理

在显著性检验中，否定或接受无效假设的依据是“小概率事件实际不可能性原理”。

小概率事件在一次试验中，几乎是不会发生的。若根据一定的假设条件计算出来该事件发生的概率很小，而在一次试验中竟然发生了，则可以认为假设的条件不正确，因此，否定假设。

（二）显著水平(Significance level）

- 用来确定否定或接受无效假设的概率标准叫**显著水平**，记作α。
- **α越小，显著性水平越高**，在生物学研究中常取**α=0.05 或 α=0.01**。
- α=0.05称为5%显著水平；α=0.01称为1%显著水平或极显著水平。

如何选择合适的α值

- 若　一个试验耗费大，可靠性要求高，不允许反复，那么α值应取小些；当一个试验结论的使用事关重大，容易产生严重后果，如药物的毒性试验，α值亦应取小些。
- 对于一些试验条件不易控制，试验误差较大的试验，可将α值放宽到0.1，甚至放宽到0.25。
-

小结：

因为显著性检验是根据“小概率事件实际不可能性原理”来否定或接受无效假设的，所以不论是接受还是否定无效假设，都没有100%的把握。

若经检验“差异显著”，对此结论有95%的把握，同时要冒5%下错结论的风险；

“差异极显著”，对此结论有99%的把握，同时要冒1%下错结论的风险；

“差异不显著”，是指在本次试验条件下，无效假设未被否定。

“差异不显著”是不是“没有差异”？（讨论）

有两种可能：

❶ 两个样本所在的总体确实没有显著差异；

❷ 两个样本所在总体平均数有差异而是因为试验误差大被掩盖了。

因而不能仅凭统计推断就作出绝对肯定或绝对否定的结论。“有很大的可靠性，但有一定的错误率”，这是统计推断的基本特点。

（三） 两类错误

❶ **I型错误 (type I error)**

第一类错误是真实情况为H_0成立，却否定了它，犯了“**弃真**”**错误**。 犯Ⅰ型错误的概率不会超过α， Ⅰ型错误也叫α错误。

❷ **II型错误(type II error)**

第二类错误是H_0实际不成立，却接受了它，犯了“纳伪”错误。 犯Ⅱ型错误的概率记为β。Ⅱ型错误又叫β错误。

第一类错误只有在否定H_0时才会发生。

犯第一类错误的概率用α表示。假设检验时，研究者可根据不同研究目的来确定α值的大小。如若取α=0.05，当拒绝H_0时，理论上犯第一类错误的概率不超过5%。

第二类错误只有在接受H_0时才会发生。

犯第二类错误的概率用β表示。β值的大小在进行假设检验时一般不知道。

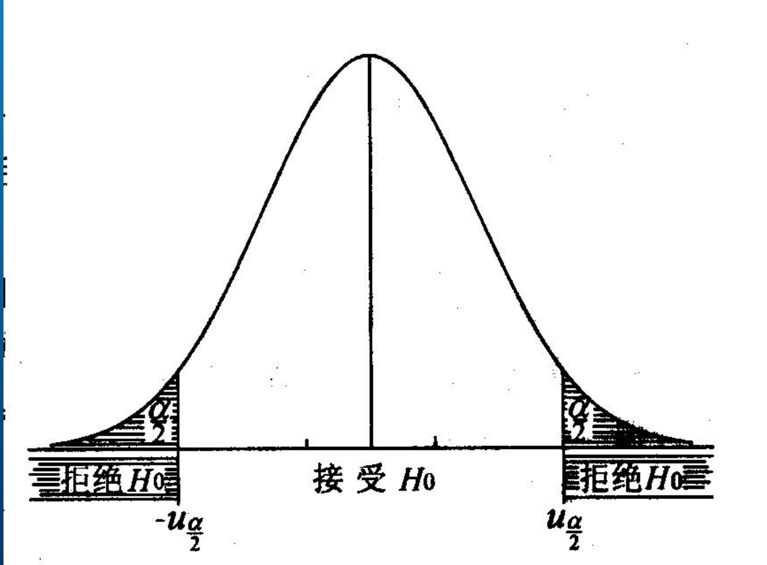
四、双侧检验与单侧检验

（一）双侧检验 (two-sided test)

$H_o: \mu_1=\mu_2$；$H_A: \mu_1\neq\mu_2$

或 $H_o: \mu=\mu_0$；$H_A: \mu\neq\mu_0$

目的在于判断有无差异，不考虑谁大谁小.

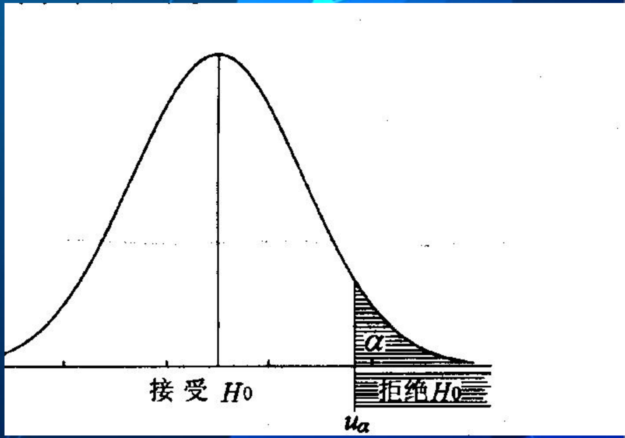


此时，在α水平上**否定域**为$(-\infty, -u_{\frac{\alpha}{2}})$和$[u_{\frac{\alpha}{2}}, +\infty]$，对称地分配在u分布曲线的两侧尾部，每侧的概率为α/2，如图所示。

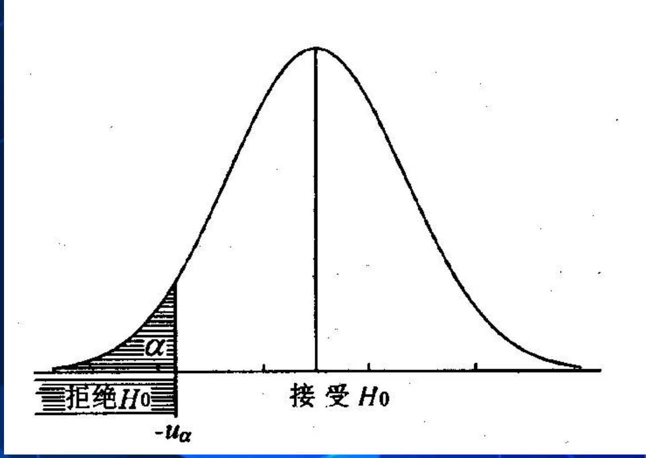
这种利用两尾概率进行的检验叫**双侧检验**，也叫**双尾检验**，$u_{\frac{\alpha}{2}}$为双侧检验的临界值。

(二）单侧检验 (one-sided test)

若无效假设为$H_0:\mu_1=\mu_2$，备择假设为$H_A:\mu_1>\mu_2$， α为右尾概率，称右侧检验，也称上尾检验，如上图所示。



若无效假设为 $H_0:\mu_1=\mu_2$，备择假设为 $H_A:\mu_1<\mu_2$， α为左尾概率，称为左尾检验，也称下尾检验，如下图所示。



这种利用一尾概率进行的检验叫**单侧检验**也叫**单尾检验**。

(三)单侧检验与双侧检验的关系

单侧检验的u_{α}=双侧检验的$u_{2\alpha}$

双侧检验显著，单侧检验一定显著；反之，单侧检验显著，双侧检验未必显著。

- (四)应用
- 选用单侧检验还是双侧检验应根据专业知识及问题的要求（分析的目的）在试验设计时就确定。
- 一般若事先不知道所比较的两个处理效果谁好谁坏，分析的目的在于推断两个处理效果有无差别，则选用双侧检验；
- 若根据理论知识或实践经验判断甲处理的效果不会比乙处理的效果差(或相反)，分析的目的在于推断甲处理是否比乙处理好(或差)，则用单侧检验。
- 一般情况下，如不作特殊说明均指双侧检验。

五、显著性检验的基本步骤？（讨论）

（一）首先对试验样本所在的**总体**作假设。

零假设

$H_0:\mu=\mu_0$ 或 $H_0:\mu-\mu_0=0$

无效假设是假设检验的基础，是将被检验的假设，它有以下三种可能来源：

- 凭以往经验或某些试验结果来设定 μ_0；
- 根据某种理论计算出 μ_0 应等于多少；
- 实际问题要求 μ_0 等于多少.

备择假设（alternative hypothesis）：

相应于H_0:$\mu=\mu_0$，则H_A有三种：

$$H_A: \mu \neq \mu_0$$

$$H_A: \mu > \mu_0$$

$$H_A: \mu < \mu_0$$

Notice: 在有专业知识可依据的情况下，应优先选用单侧检验，因为单侧检验建立在另一侧实际不可能的基础上，可提高检验精度.

（二）选择合适的显著水平

根据不同的试验要求选取不同的α值，一般常用的为0.05和0.01.

计算出的概率大于0.05，称之为“没有显著差异”；

计算出的概率小于0.05，称之为“差异显著”；（再作进一步比较）

计算出的概率小于0.01，称之为“差异极显著”.

（三）选择合适的统计量，并研究试验所得统计量的抽样分布。

根据不同的目的采用不同的检验方法：对平均数做检验，u检验（σ^2已知）或t检验（σ^2未知），单个样本方差检验用 χ^2 检验，两个样本方差用F检验.

u, t, χ^2，F 称为检验统计量.

（四）建立H_0的拒绝域，查表确定临界值。

- 根据备择假设，建立相应的H_0的拒绝域

（五）做出推断及生物学解释。

$P \geq 0.05$　不认为是小概率事件，接受H_0

$P<0.01$或$P<0.05$为小概率事件，则

否定$H_0:\mu=\mu_0$；接受H_A

根据结果对原问题做出明确、合理的解释。

小结：显著性检验中应注意的问题

① 试验之前进行严格合理的试验设计或抽样设计；

② 根据不同的试验设计方法，选择不同的显著性检验方法；

③ 要正确理解“差异不显著，差异显著和差异极显著”的统计意义；

显著水平的高低只表示下结论的可靠性程度的高低；

显著性检验只是用来确定无效假设能否被推翻，而不能证明无效假设是否正确.

④　合理建立统计假设，正确计算统计量；

⑤　统计结论不能绝对化，统计分析应与专业知识相结合；

⑥　报告结论时应列出检验用的统计量的值、P值范围、注明是双侧还是单侧检验.