《数理统计》

2021年4月2日 21:39

https://zhuanlan.zhihu.com/p/31743949

大数定律及中心极限定理

a. **依概率收敛，切比雪夫不等式**

- 依概率收敛
  - $$\lim_{n\to+\infty} P\{|\frac{n_A}{n}-p|\geq\epsilon\}=0$$
  - 记为 $Y_n \xrightarrow{p} c$ 当 $n\to+\infty$
- 依概率收敛的性质
  - 若$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$ 且 $g(x,y)$ 在点 a,b 处连续
  - $X_n \xrightarrow{p} g(a,b)$ 当 $n\to+\infty$ 时
  - 如，$X_n+Y_n \xrightarrow{P} a+b$ 当 $n\to+\infty$
  - 特别的 $f(X_n)\xrightarrow{P} f(a)$
- 切比雪夫不等式 Chebyshev's inequality

  $$P\{|X-\mu|\geq\epsilon\}\leq\frac{\sigma^2}{\epsilon^2}$$
- 等价形式

  $$P\{|X-\mu|<\epsilon\}\geq 1-\frac{\sigma^2}{\epsilon^2}$$
- 切比雪夫不等式的性质

  适用范围： 期望、方差存在的随机变量

  重要性： 可以对于随机变量落在期望附近的区域内或外给出一个界的估计

  切比雪夫不等式应用范围广，但结果比较粗糙

b. **大数定律**

- 频率的稳定值记为概率，这个结论可以用”大数定律“来描述
- 伯努利大数定律
  - $$\lim_{n\to+\infty} P\{|\frac{n_A}{n}-p|\epsilon\}=0$$
  - 即

    $$\frac{n_A}{n}\xrightarrow{p} p$$
  - 意义：提供了通过实验来确定事件概率的方法
- 大数定律
  - 设 $X_1,X_2,\ldots,X_n,\ldots$ 是一列随机变量，则在一定条件下，随机变量序列 $Y_n=\frac{X_1+\ldots+X_n}{n}$ 收敛到 μ，当 $n\to+\infty$
  - 含义：依概率收敛
  - 当 X_i 期望相同时，$\mu=E(X_i)$

- 切比雪夫大数定律的推论
  - $X_1, X_2, \ldots, X_n, \ldots$ 为相互独立的随机变量，且具有相同的期望 μ，相同的方差σ^2，那么
$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$
，当 $n \to +\infty$
- 辛钦大数定律
  - $X_1, X_2, \ldots, X_n, \ldots$ 为相互独立同分布的随机变量，且期望 μ 存在，那么
$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu$$
，当 $n \to +\infty$

c. 中心极限定理
- 有许多随机变量，它们是由大量的相互独立的随机变量的综合影响所形成的，而其中每个个别因素的作用都很小，这种随机变量往往服从或近似服从正态分布，或者说他们的极限是正态分布，中心极限定理正是数学上论证了这一现象，它在长达两个世纪的时期内曾是概率论研究的中心课题
- 独立同分布的中心极限定理（CLT）
  - 设 $X_1, X_2, \ldots, X_n, \ldots$ 相互独立且同分布，$E(X_i) = \mu, D(X_i) = \sigma^2, i = 1, 2, \ldots$ 则对于充分大 n 的，有
  - $$\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)$$
  - 此时
  - $$P(a < \sum_{i=1}^{n} X_i \le b) \approx \Phi(\frac{b - n\mu}{\sqrt{n}\sigma}) - \Phi(\frac{a - n\mu}{\sqrt{n}\sigma})$$
- 注意，CLT仅仅是分布类型上的一种近似
- 德莫弗-拉普拉斯定理
  - 即二项分布可以用正态分布逼近
  - $$n_A \overset{\text{近似}}{\sim} N(np, np(1-p))$$

统计量与抽样分布

a. 总体，样本
- 数理统计是一门以数据为基础的学科。数理统计学的任务就是如何获得样本和利用样本，从而对事物的某些未知方面进行分析、推断并做出一定的决策
- 提示：若非数学相关专业，从本章开始虽然难度增高，但要求会降低很多，实际考试难度是降低的
- 总体

- 研究对象的全体
- 总体的某个指标 X 可以看成一个随机变量
- 有时也直接称 X 为总体
- X 具有分布函数 $F(x)$

- 样本
  被抽取的部分个体
- 简单随机样本
  - 随机样本 $(X_1, X_2, \ldots, X_n)$ 中，每个 X_i 与 X_n 是相互独立的随机变量
  - 这些样本和总体 X 同分布
- 获得简单随机样本
  简单随机抽样
  对于有限个体采用放回抽样
  对于无限总体（或很大的总体）采用不放回抽样
- 注意
  - 对样本进行一次观测，得到实际数值 $x_1, x_2, \ldots, x_n$ 成为样本观测值
  - 一般情形下，两次观测，样本值是不同的

b. 统计量，常用统计量

- 统计量
  统计量是样本的不含任何位置参数的函数
- 常用统计量
  - 样本均值
  - $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
  - 样本方差
  - $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
  - 拓展：知乎问题：为什么样本方差（sample variance）的分母是 n-1?
  - 亦可参考本章末尾的作者拓展
  - 样本矩
  - k 阶矩
    - $$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$
  - k 阶中心矩
    - $$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

https://www.zhihu.com/question/20099757/answer/26586088

c. X^2分布

- 读作" 卡方分布 "，对应希腊字母里的 'chi'
- 设随机变量 $X_1, \dots, X_n$ **相互独立**且都服从 $N(0,1)$
  - $$\chi^2 = \sum_{i=1}^{n} X_i^2$$
  - 记为 $\chi^2 \sim \chi^2(n)$，n 为自由度
- 概率密度（不重要）
  - $f_n(x) = \begin{cases} \dfrac{1}{2\Gamma(n/2)} \left(\dfrac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$
  - 其中

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

- 性质
  - 设 $\chi^2 \sim \chi^2(n)$，则 $E(\chi^2) = n, D(\chi^2) = 2n$
  - 提示：正态分布有如下性质：若 $X \sim N(\mu, \sigma^2)$，则$X^2 \sim N(\sigma^2 + \mu^2, \sigma^2(2\mu^2 - \sigma^2))$，该性质证明较复杂，在此不做证明
  - χ^2 分布的可加性
  - 设 $Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$ 且它们相互独立，则
  - $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$
  - 以上结论也可以推广到有限个随机变量的情形
- 上 α 分位数
  - 给定 $\alpha, 0 < \alpha < 1$，称满足条件 $P(\chi^2 > \chi^2_\alpha(n) = \alpha)$ 的点 $\chi^2_\alpha(n)$ 为 $\chi^2(n)$ 分布的上 α 分位数
  - 注意，$\chi^2_\alpha(n)$ 是一个实数
  - 该值可以查表或通过计算机求解

d. t分布， F分布

- t分布

- 设 $X \sim N(0,1), Y \sim \chi^2(n)$，且 X 和 Y 相互独立，则
- $$T = \frac{X}{\sqrt{(Y/n)}}$$
- 称为服从自由度 n 的 t 分布，也称学生氏分布，记为
- $$T \sim t(n)$$
- 概率密度（不重要）
- $$f(x;n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < +\infty$$
- 特别的，$n=1$ 的 t 分布就是柯西分布
- $$f(x;1) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < +\infty$$
- 当 $n \to +\infty$ 时的 t 分布就是标准正态分布
- 上 α 分位数
- 满足条件$P(t > t_\alpha(n) = \alpha)$ 的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位数
- 由于 t 分布是关于 0 对称的，则 $t_{1-\alpha}(n) = -t_\alpha(n)$

- F分布
  - 设 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$，且 X 和 Y 相互独立，则
  - $$F = \frac{X/n_1}{Y/n_2}$$
  - 称为服从自由度 (n_1, n_2) 的 F 分布，记为
  - $$F \sim F(n_1, n_2)$$
  - 其中 n_1 为第一自由度，n_2 为第二自由度
  - 性质
  - 若 $F \sim F(n_1, n_2)$，则

  $$\frac{1}{F} \sim F(n_2, n_1)$$

  - 概率密度（不重要）
  - $$f(x;n_1,n_2) = \begin{cases} \frac{1}{B(n_1/2, n_2/2)} n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} x^{\frac{n_1}{2}-1}(n_2+n_1x)^{\frac{n_1+n_2}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$
  - 其中，

  $$B(a,b) = \int_0^1 x^{\alpha-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

  - 上 α 分位数
  - 满足条件$P(F > F_\alpha(n_1,n_2) = \alpha)$ 的点 $F_\alpha(n_1,n_2)$ 为 $F(n_1,n_2)$ 分布的上 α 分位数
- 实例

  实例，设 X, Y, Z 相互独立，均服从 $N(0,1)$，则

实例，设 X, Y, Z 相互独立，均服从 $N(0,1)$，则

- $$X^2+Y^2+Z^2 \sim \chi^2(3)$$
- $$\frac{X}{\sqrt{(Y^2+Z^2)/2}} \sim t(2)$$
- $$\frac{2X^2}{Y^2+Z^2} \sim F(1,2)$$
- 若 $t \sim t(n)$，则 $t^2 \sim F(1,n)$

e. 单个正态总体的抽样分布

- 定理一
  - 设总体 $X \sim N(\mu, \sigma^2)$ 样本均值

    $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

    样本方差

    $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

    则
  - $$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$
  - $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
  - 证明过程略
- 定理二
  - 背景：由定理一知，

    $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
  - 上式的意义：$\bar{X}$ 标准化后服从标准正态分布
  - 当 σ 未知时，可用 S^2 来替代
  - 则引出定理二：

    $$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$
  - 提示：该定理可作为第七章提到的枢轴量
  - 拓展：在概率论和统计学中，学生t-分布（Student's t-distribution），可简称为 t 分布，用于根据小样本来估计呈正态分布且方差未知的总体的均值。如果总体方差已知（例如在样本数量足够多时），则应该用正态分布来估计总体均值。

  实例：
  - $E(S^2) = \sigma^2$
  - $$D(S^2) = \frac{2\sigma^4}{}$$

- $E(S^2) = \sigma^2$
- $$D(S^2) = \frac{2\sigma^4}{n-1}$$
- 上式说明，随着样本量的增大，样本均值的偏差 $D(S^2)$ 减小

提示：这一章的内容看似难度陡增，但实际上知识点都是环环相扣的，概率论的各种公式实际上都是出的。明白了这一点对一些看似莫名其妙的公式就无需害怕了。比如科学家发现在统计中，样本较小

f. **两个正态总体的抽样分布**

- 定理三
  - 设样本 $(X_1, \ldots, X_{n_1})$ 和 $(Y_1, \ldots, Y_{n_1})$ 分别来自总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$ 并且它们相互独立，则

1. $$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2} \Big/ \frac{\sigma_1^2}{\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$
2. $$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$
3. 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 时
   - $$\frac{(\bar{X} - \bar{Y}) - \mu_1 - \mu_2}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$
   - 其中
   - $$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

提示：上一小节和这一小节的内容主要是为了后续的区间估计和假设检验做论证

参数估计

a. **矩估计**

- 参数估计的形式

  点估计

  实例：明天的最高温度为12度

  区间估计

  实例：明天的最高温度在12度-13度之间
- 常用的点估计方法

  矩估计法

  极大似然估计法
- 矩估计法

  以样本矩估计总体局，以样本矩的函数估计总体矩的函数

  其实就是解方程，有几个未知参数就用到第几阶矩，其中未知量是参数，已知量是矩
- 实例

- $X \sim U(a,b)$ 求 a,b 的矩估计量
- $$\mu_1 = E(X) = \frac{a+b}{2}, \nu_2 = \frac{b-a}{12}$$
- 解得 $\hat{a} = \bar{X} - \sqrt{3B_2}, \hat{b} = \bar{X} + \sqrt{3B_2}$
- 提示：其中的 μ_1 代表一阶原点矩，ν_2 代表二阶中心矩，$\bar{X}$ 代表样本一阶原点矩（均值），B_2 代表样本二阶中心矩（方差）

- 矩估计有随意性，估计方式符合实际即可。但注意，如估计正态分布的方差时，
  方差的矩估计不是样本方差，而是样本的二阶中心矩

b. 极大似然估计

- 似然函数
  - 离散型
  - $$L(\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$
  - 连续型
  - $$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$
- 极大似然原理
  - $$L(\hat{\theta}(x_1,\ldots,x_n)) = \max_{\theta\in\Theta} L(\theta)$$
  - 即令似然函数最大
- 求最大值时可以转为求 $\ln L(\theta)$ 的最大值，求极值时也可以利用导数。若似然函数是单调的，则根据实际情况取值
- 实例
  - $X \sim N(\mu,\sigma^2), \quad \hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = B_2$
  - 即正态分布的矩估计和极大似然估计结果完全一样
  - $X \sim U(a,b), \quad \hat{a} = \min\{X_1,\ldots,X_n\}, \quad \hat{b} = \max\{X_1,\ldots,X_n\}$
- 注意：矩估计和极大似然估计是两种完全不同的估计方法，极大似然估计需要知道分布，而矩估计不需要知道

c. 估计量的评价准则，无偏性

- 常用的评价准则
  无偏性准则
  有效性准则
  均方误差准则
  相合性准则
- 无偏性
  - 即 $E(\hat{\theta}) = \theta$
  - 即没有系统误差
- 渐进无偏
  $$\lim_{n\to+\infty} E(\hat{\theta}) = \theta$$
- 实例
  - 设 $E(X) = \mu, \quad D(X) = \sigma^2$

- 设 $E(X)=\mu,\quad D(X)=\sigma^2$
- $$E(B_2)=\frac{n-1}{n}\sigma^2$$
- $\bar{X},S^2$ 是无偏估计，B_2 是渐进无偏估计
- 设 $X\sim U(0,\theta)$
- 矩估计无偏
- 极大似然估计有偏

- 纠偏方法
  - 如果有偏结果和真正的结果之间有函数关系，则使用反函数即可纠偏
  - 这也是第六章中提及的样本方差前面是
    $$\frac{1}{n-1}$$
    的原因

d. 有效性，均方误差

- 有效性准则
  - 设 $\hat{\theta}_1,\hat{\theta}_2$ 是 θ 的两个无偏估计。如果 $D(\hat{\theta}_1)\le D(\hat{\theta}_2)$ 对一切 $\theta\in\Theta$ 成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效
- 均方误差准则
  - $E(\hat{\theta}-\theta)^2$
  - 记为 $Mse(\hat{\theta})$
  - 注意：若 $\hat{\theta}$ 是 θ 的无偏估计，则有 $Mse(\hat{\theta})=D(\hat{\theta})$

e. 相合性

- 相合性准则
  - $\hat{\theta}_n\xrightarrow{P}\theta$ （依概率收敛）
  - 则称 $\hat{\theta}_n$ 是 θ 的相合估计量或一致估计量
- 实例
  - 随机取一个样本作为期望的估计不是相合估计
  - 设总体 $X\sim Y[0,\theta]$，证明

f. 置信区间，置信限

- 置信区间
  - 若有两个统计量 $\hat{\theta}_L=\hat{\theta}_L(X_1,\dots,X_n),\quad \hat{\theta}_U=\hat{\theta}_U(X_1,\dots,X_n)$ 使得 $P\{\hat{\theta}_L(X_1,\dots,X_n)<\theta<\hat{\theta}_U(X_1,\dots,X_2)\}\ge 1-\alpha$ 则称 $(\hat{\theta}_L,\hat{\theta}_U)$ 为 θ 的置信水平为 $1-\alpha$ 的双侧置信区间。
- 置信区间的意义
  - 反复抽样多次，每个样本值确定一个置信区间，每个这样的区间包含 θ 真值的比例约为 $1-\alpha$
  - 提示：对于一个具体的区间而言，或者包含真值，或者不包含真值，无概率可言。
- 单侧置信区间
  - 如果 $P\{\hat{\theta}_L(X_1,\dots,X_n)<\theta\}\ge 1-\alpha$ 则称 $\hat{\theta}_L$ 为单侧置信下限，反之为置信上限
- 单侧置信区间和双侧置信区间的关系
  - 设 $\hat{\theta}_L$ 是 θ 的置信水平为 $1-\alpha_1$ 的单侧置信下限
  - 设 $\hat{\theta}_U$ 是 θ 的置信水平为 $1-\alpha_2$ 的单侧置信下限
  - 则 $(\hat{\theta}_L,\hat{\theta}_U)$ 是 θ 的置信度为 $1-\alpha_1-\alpha_2$ 的双侧置信区间

- 设 $\hat{\theta}_L$ 是 θ 的置信水平为 $1-\alpha_1$ 的单侧置信下限
- 设 $\hat{\theta}_U$ 是 θ 的置信水平为 $1-\alpha_2$ 的单侧置信下限
- 则 $(\hat{\theta}_L, \hat{\theta}_U)$ 是 θ 的置信度为 $1-\alpha_1-\alpha_2$ 的双侧置信区间

- 精确度和误差限
  - 称置信区间的平均长度 $E(\hat{\theta}_U - \hat{\theta}_L)$ 为精确度。在给定的样本容量下，置信水平越高，精确度越低。精确度高，置信水平越低
- Neyman 原则
  - 在置信水平相同的情况下选择精确度尽可能高的置信区间

g. 枢轴量法

- 枢轴量法
  - $\hat{\theta}_L < \theta < \hat{\theta}_U$
  - G 的分布是已知的且不依赖于任何未知参数，也不依赖于总体的分布。称 G 是枢轴量
  - 直观理解：理解枢轴量可以抛开总体，它应用于估计某个参数之上，因此只需要考虑估计正确的可能性。它和总体没有关系，把待估参数 θ 理解为一个单独的随机变量即可。
- 正态总体下常见的枢轴量
  - **提示：以下公式的证明在第六章后两个小节**
  - 单个正态总体
  - μ （σ^2 已知）
    - $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
  - μ （σ^2 未知）
    - $$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$
  - σ^2 （μ 未知）
    - $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
  - 二个正态总体
  - $\mu_1-\mu_2$ （σ_1^2, σ_2^2 已知）
    - $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$
  - $\mu_1-\mu_2$ （$\sigma_1^2 = \sigma_2^2$ 且未知）
    - $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{2}{n_2}}} \sim t(n_1+n_2-2)$$
    - 其中

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

  - σ_1^2/σ_2^2 （μ_1, μ_2 未知）

- σ_1^2/σ_2^2（μ_1,μ_2 未知）
  - $$\frac{S_1^2}{S_2^2} \bigg/ \frac{\sigma_1^2}{\sigma_2^2} \sim F(n_1-1,n_2-1)$$

h. 单个正态总体均值的区间估计

- σ^2 已知
  - $$(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2})$$
  - 单侧置信下限为
    $$\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha}$$
  - 单侧置信上限为
    $$\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha}$$
- σ^2 未知
  - $$(\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1))<\mu<\bar{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)$$
  - 单侧置信下限为
    $$\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha}(n-1)$$
  - 单侧置信上限为
    $$\bar{X}+\frac{S}{\sqrt{n}}t_{\alpha}(n-1)$$
- 非正态总体均值
  - 当 n 充分大（一般为 $n>30$）时，有中心极限定理
  - $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim^{近似} N(0,1)$$

i. 成对数据均值差，单个正态总体方差的区间估计

- 成对数据差
  - $D_i = X_i - Y_i$
- 估计 μ_D
  - $$(\bar{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}})$$
- 估计 σ^2
  - $$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2)}{\chi_{1-\alpha/2}(n-1)}\right)$$
  - 单侧置信下限
  - $$\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}$$
  - 单侧置信下限
  - $$\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}$$
  - 注意：上述置信区间不是最优解，但为了计算方便采用上述区间。

j. 两个正态总体参数的区间估计

- $\mu_1 - \mu_2$（σ_1^2, σ_2^2 已知）
  - $$\left((\bar{X}-\bar{Y}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$
- $\mu_1 - \mu_2$（$\sigma_1^2 = \sigma_2^2$ 且未知）
  - $$\left((\bar{X}-\bar{Y}) \pm t_{\frac{\alpha}{2}}(n_1+n_2-2)S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$
  - 其中
  $$S_W^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$
- $\mu_1 - \mu_2$（σ_1^2, σ_2^2 未知且样本量较大）
  - 可以用 S_1^2 和 S_2^2 估计 σ_1^2 和 σ_2^2
  - $$\left((\bar{X}-\bar{Y}) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$
- $\mu_1 - \mu_2$（σ_1^2, σ_2^2 未知且样本量较小）
  - $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim^{近似} t(k)$$
  - 其中 $k \approx (n_1-1, n_2-1)$
  - 则置信区间为

- 其中 $k \approx (n_1 - 1, n_2 - 1)$
- 则置信区间为
- $$\left((\bar{X} - \bar{Y}) \pm t_{\frac{\alpha}{2}}(k)\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

- σ_1^2/σ_2^2（μ_1, μ_2 未知）
  - $$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)}\right)$$
- 提示：以上求解时并没有直接求置信区间的最小值而是近似的去除两端取中间

假设检验

a. **假设检验的基本思想**
b. **单个正态总体参数假设检验（标准差已知，检验）**
c. **单个正态总体参数假设检验（标准差未知，检验）**
d. **单个正态总体参数假设检验（成对数据检验和参数的检验）**
e. **两个正态总体参数假设检验(比较两个正态总体均值的检验)**
f. **两个正态总体参数假设检验(比较两个正态总体方差的检验)**
g. **拟合优度检验**

方差分析与回归分析（略）

a. **单因素方差分析**
b. **单因素方差分析（参数估计及均值的多重比较）**
c. **回归分析（参数估计）**
d. **回归分析（模型检验与应用）**