

样本及抽样分布

讲师：Jeary

目录

01

随机样本

02

抽样分布

03

总结

目标

 通过本章课程的学习，您将能够：

- 了解随机样本的概念
- 熟悉随机抽样及抽样分布

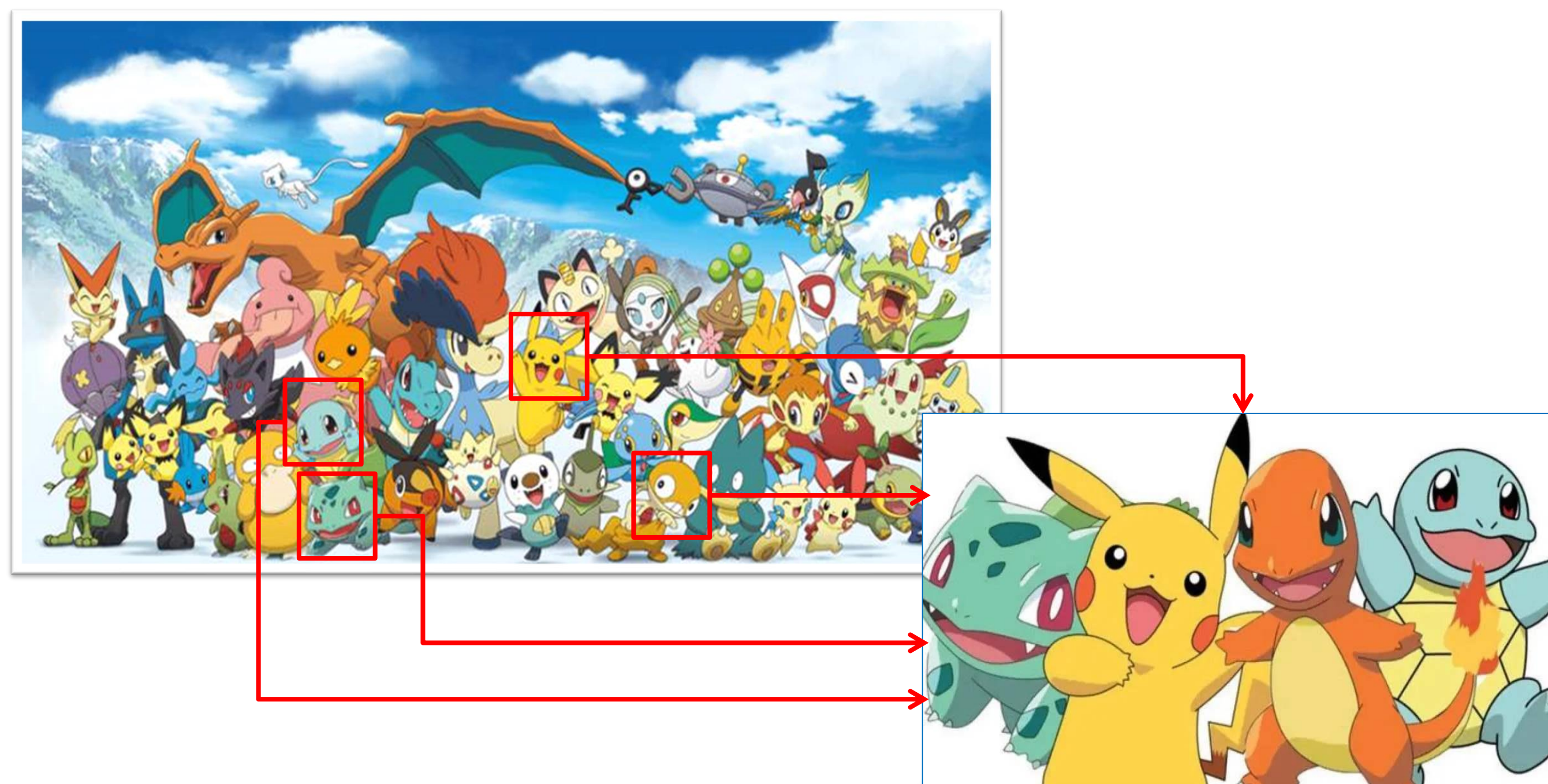


随机样本

基本概念

总体： 研究的对象的全体

个体： 总体中的每个成员



基本概念

容量

总体中包含的个体的数量

有限总体

容量为有限的总体

无限总体

容量为无限的总体

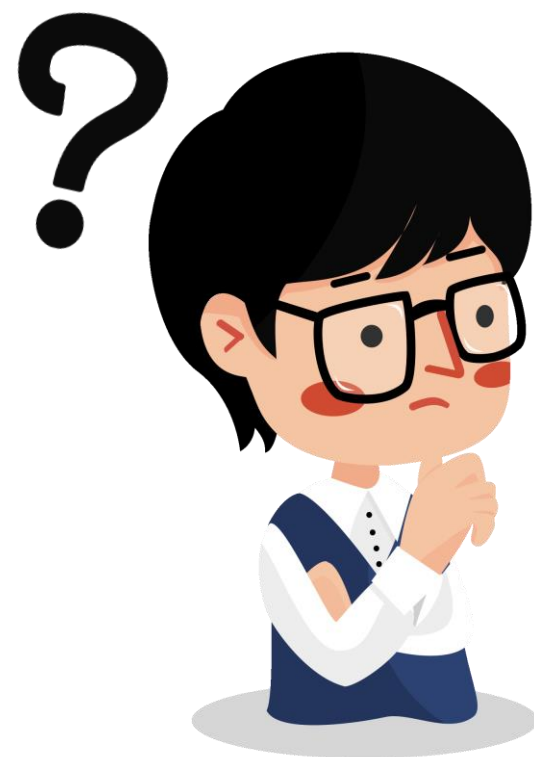
简介

条件：相同情况下，随机观察样本的值

操作：对总体反复随机独立观察

观察N次，记录： x_1, \dots, x_n ， x_1, \dots, x_n 均是随机变量， x_1, \dots, x_n 称为样本，每个 x_i 称为样本点， x_i 称为样本 X 样本值

思考题



01

我们针对报名数据分析课程的6W名学员进行阶段测评，调查下学员的学习情况，总体、个体、容量分别是什么？



抽样分布

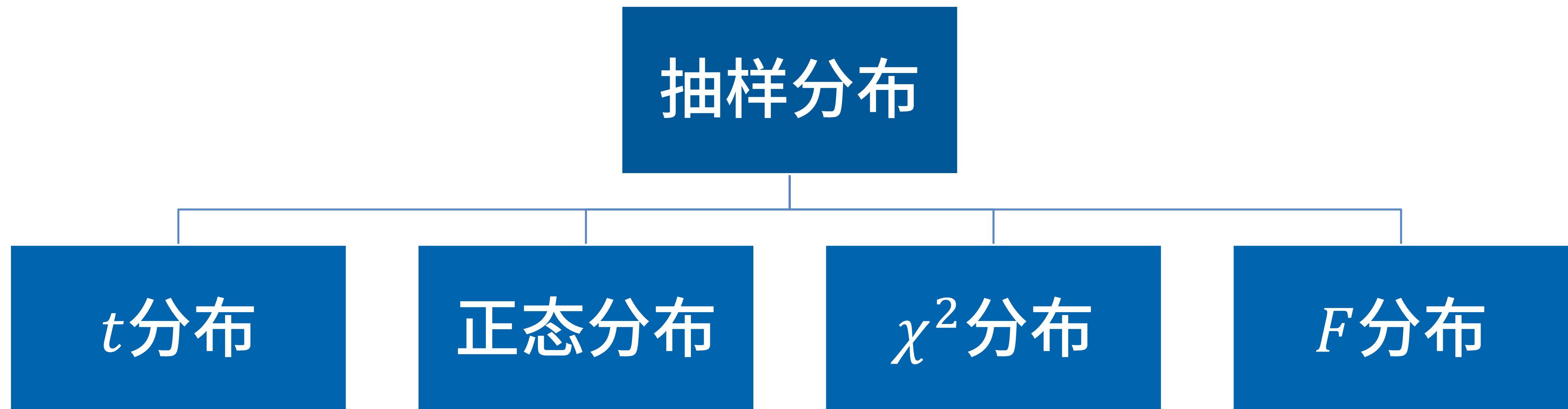
简单随机抽样

条件：随机性、独立性

适用场景：

- 有放回抽样：有限总体
- 不放回抽样：无限或者大总体

抽样分布



t分布

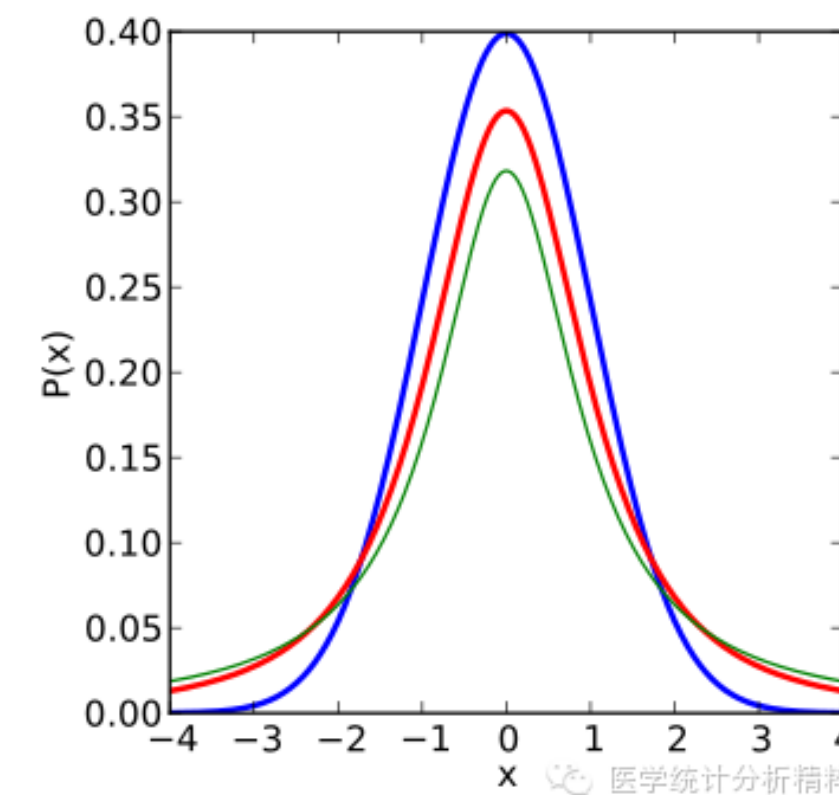
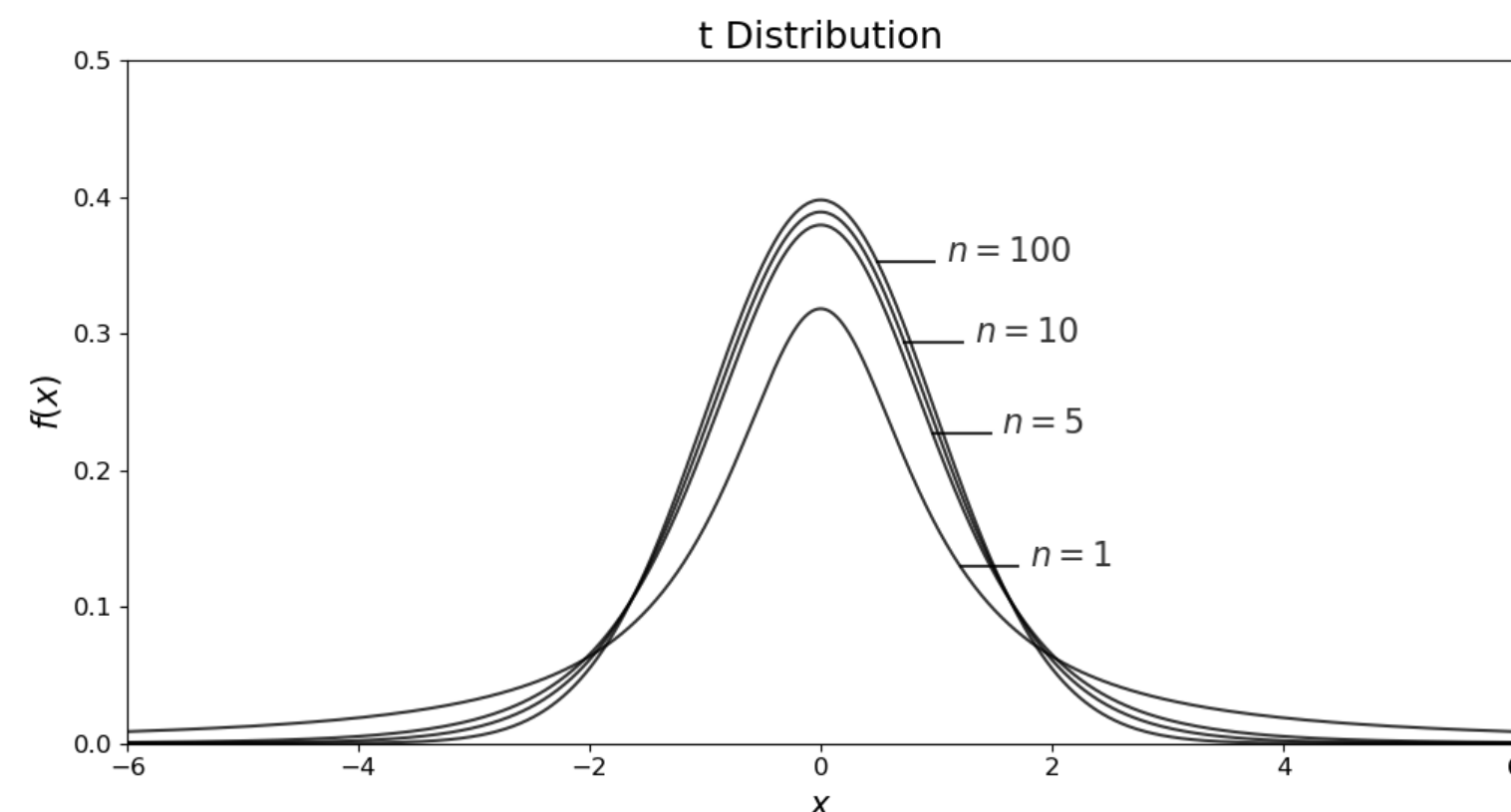
■ 定义

设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X, Y 相互独立, 则称随机变量:

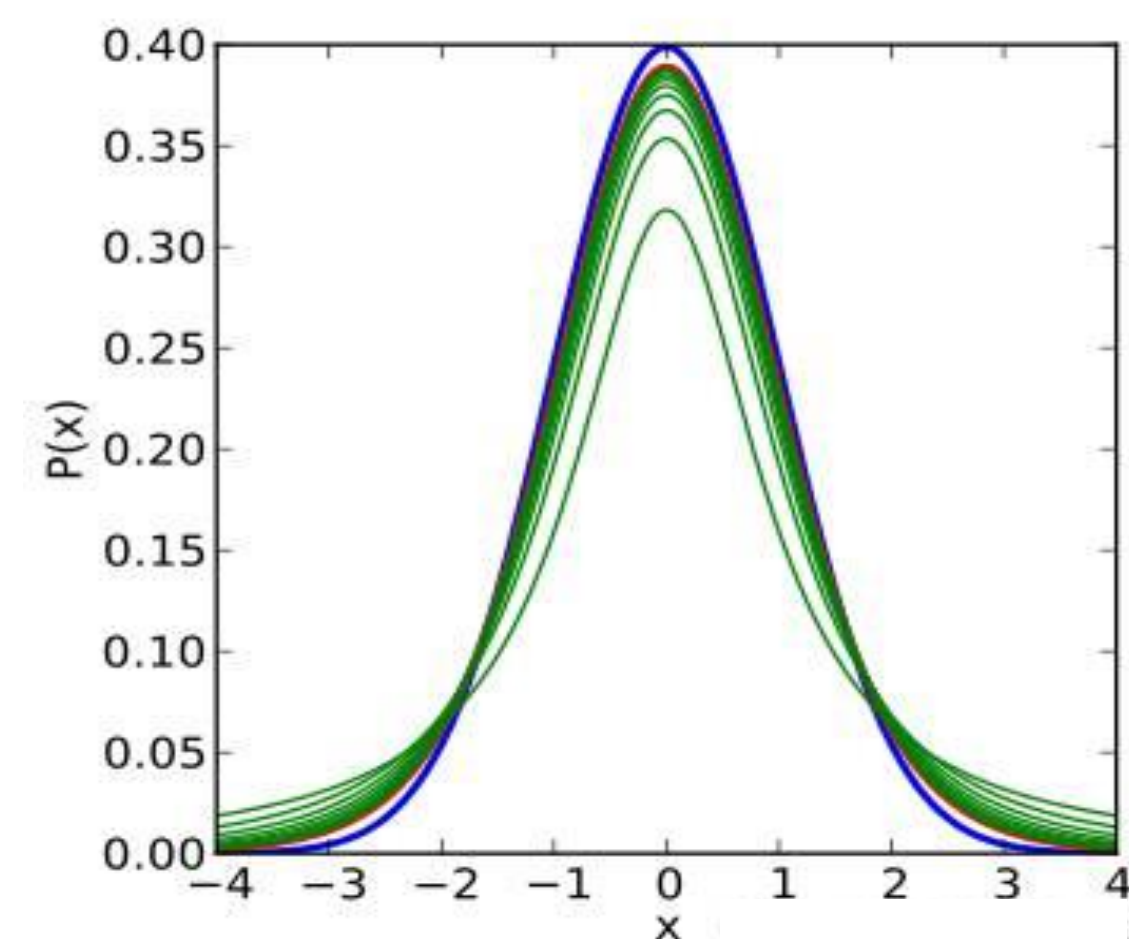
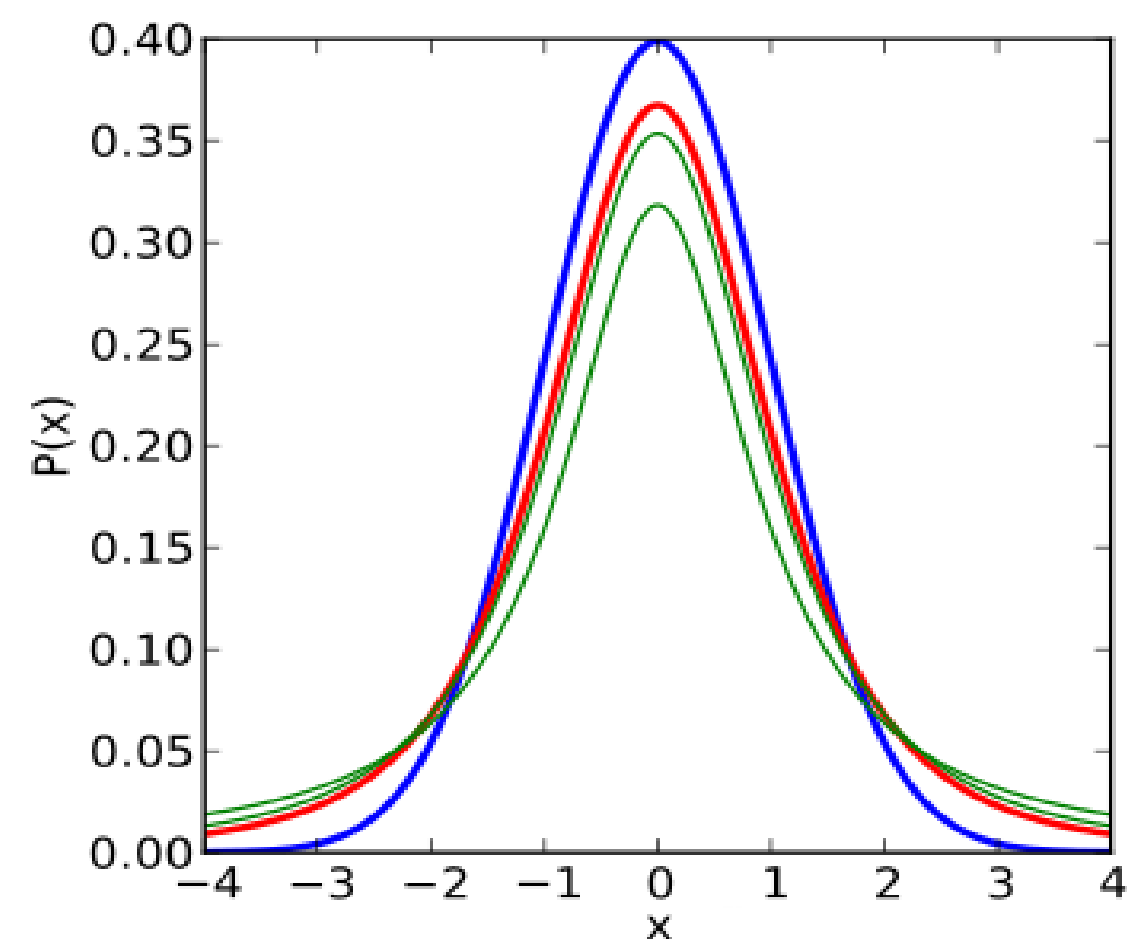
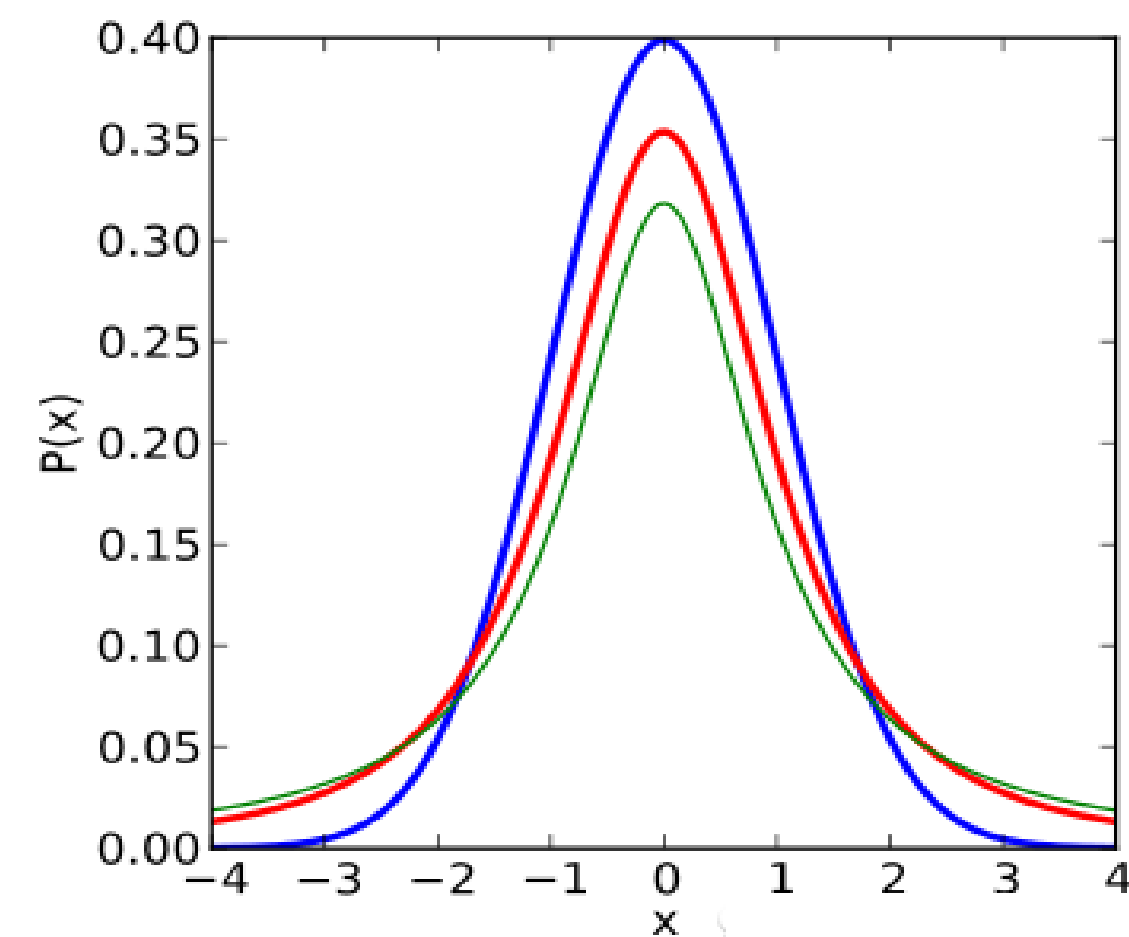
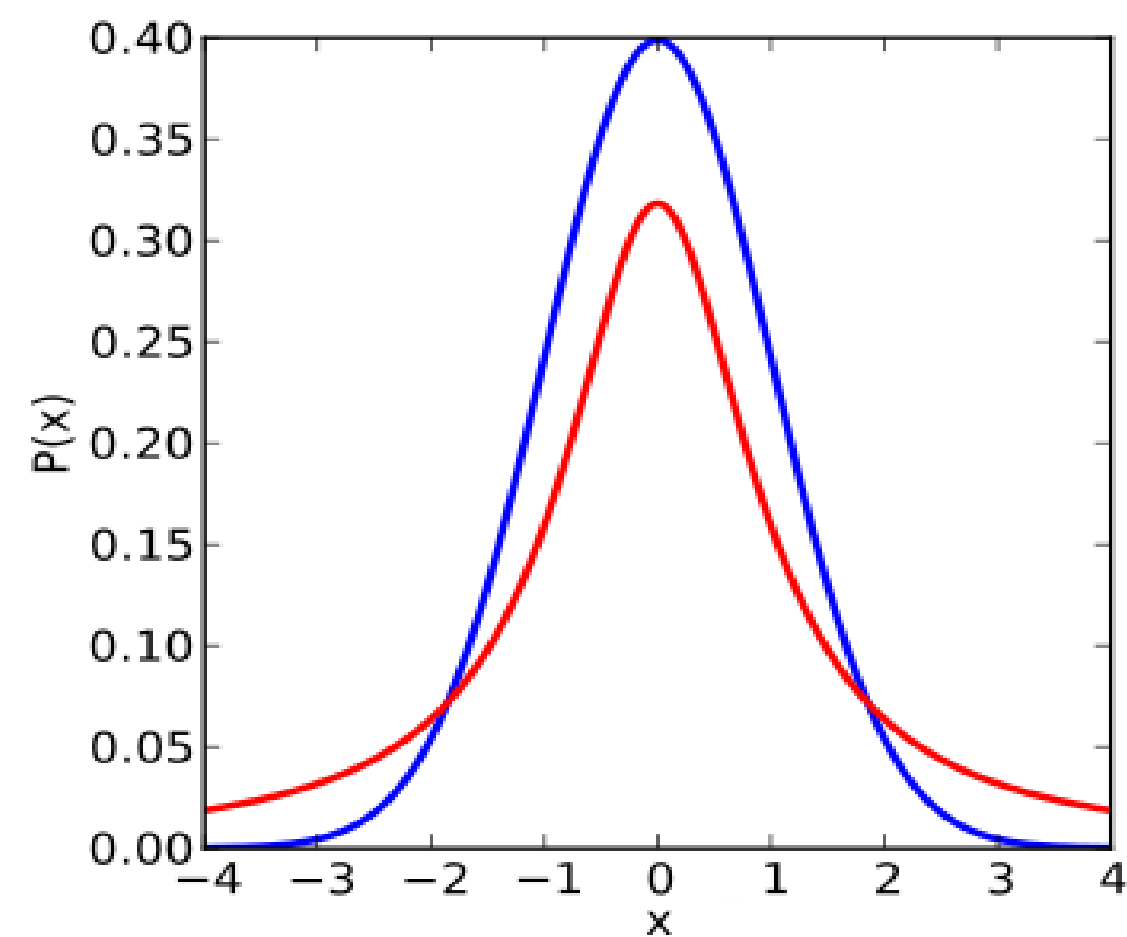
$$t = \frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$, t 分布又称学生分布

$t(n)$ 分布的概率密度函数为: $h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, -\infty < t < \infty$



T分布与正态分布的关系



t 分布的优势

研究样本量要求低

缩小标准差
容易找到p值

点估计更准确

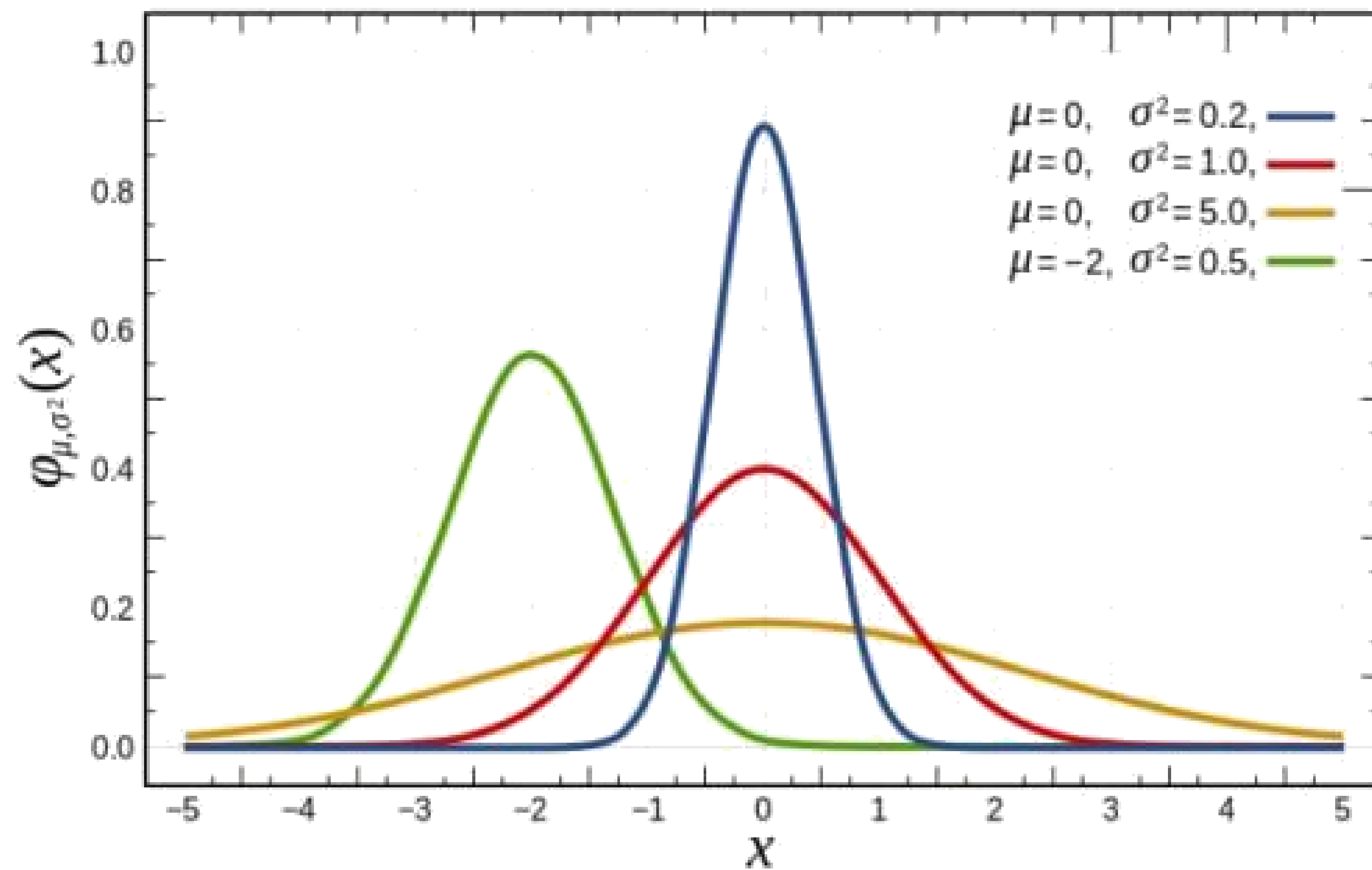
稳健回归

t 分布-实例

已知在线教育行业内学员对课程的满意度平均为3.5分，Y公司对已开课的《数据分析》学员做满意度调查得出平均分为3.71分，其满意度是否超过业内的平均分呢？



正态分布



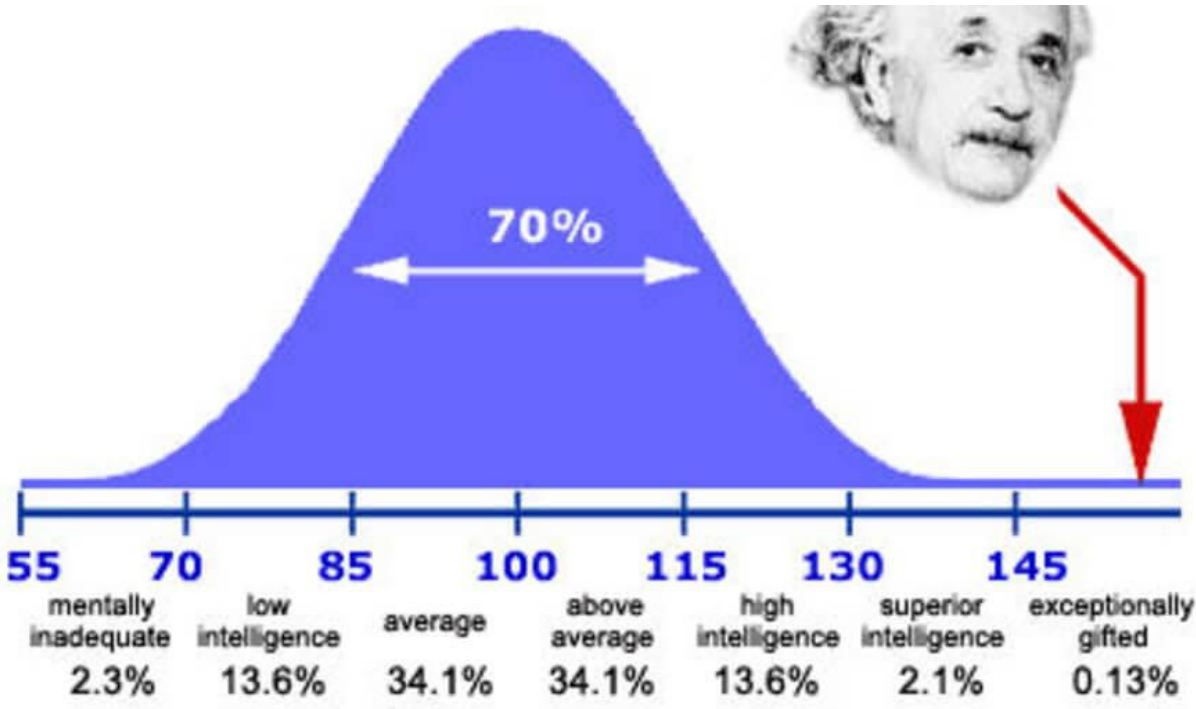
正态分布现象



员工绩效



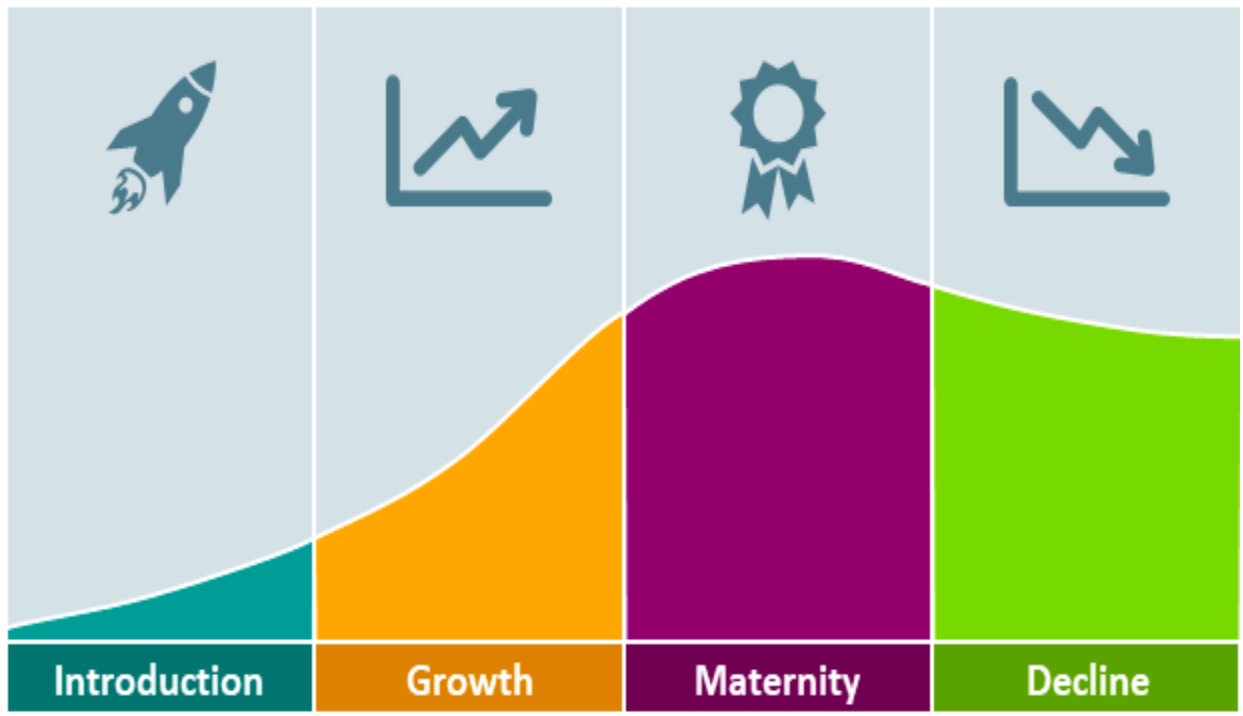
身高



智商



能力



产品的生命周期



产品质量

χ^2 分布

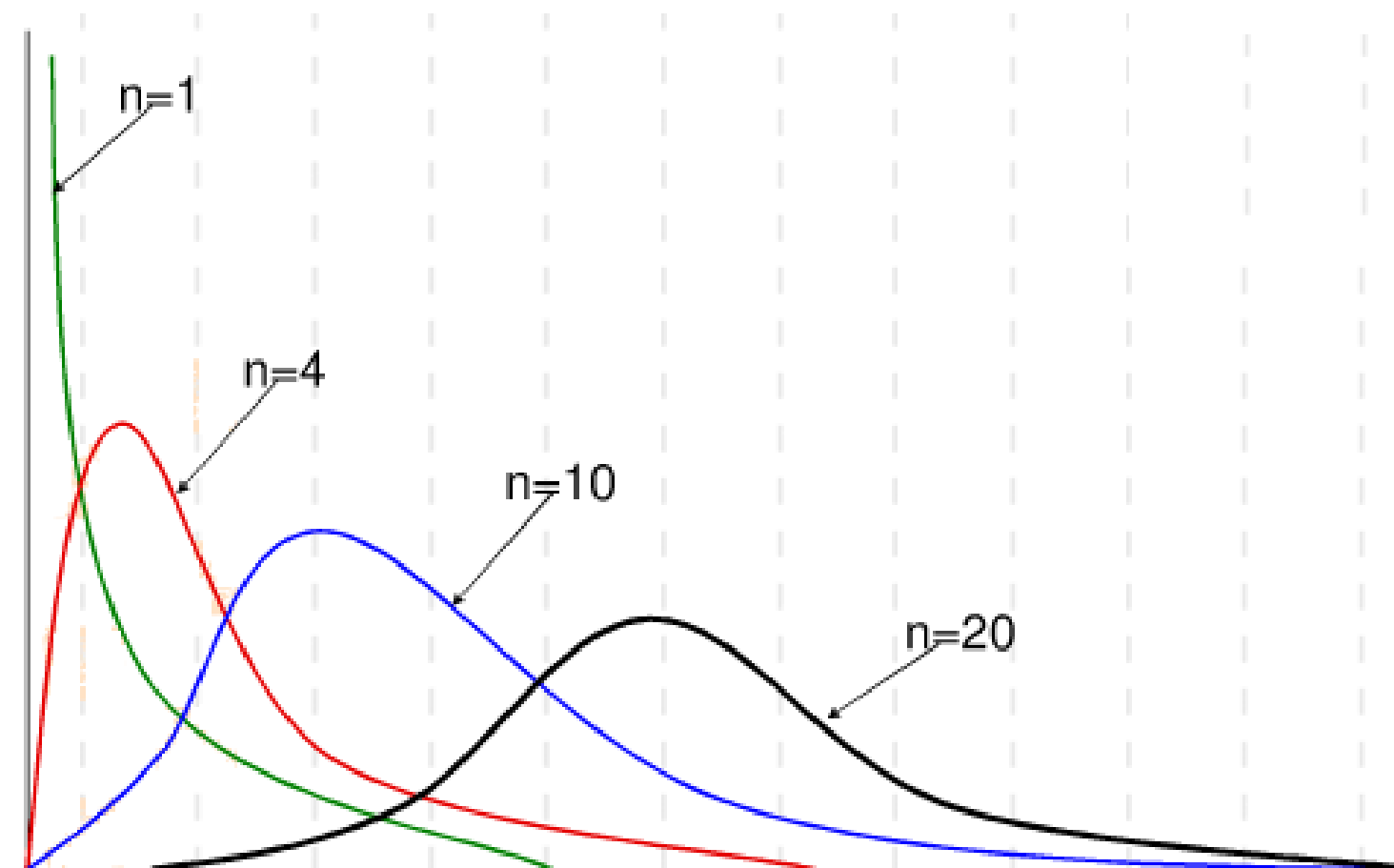
■ 定义

设 X_1, X_2, \dots, X_n 是来自总体 $N(0, 1)$ 的样本，则称统计量

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

服从自由度为 n 的 χ^2 分布，记为： $\chi^2 \sim \chi^2(n)$

自由度是指 $X_1^2, X_2^2, \dots, X_n^2$ 中包含的独立变量的个数



χ^2 分布

概率密度:

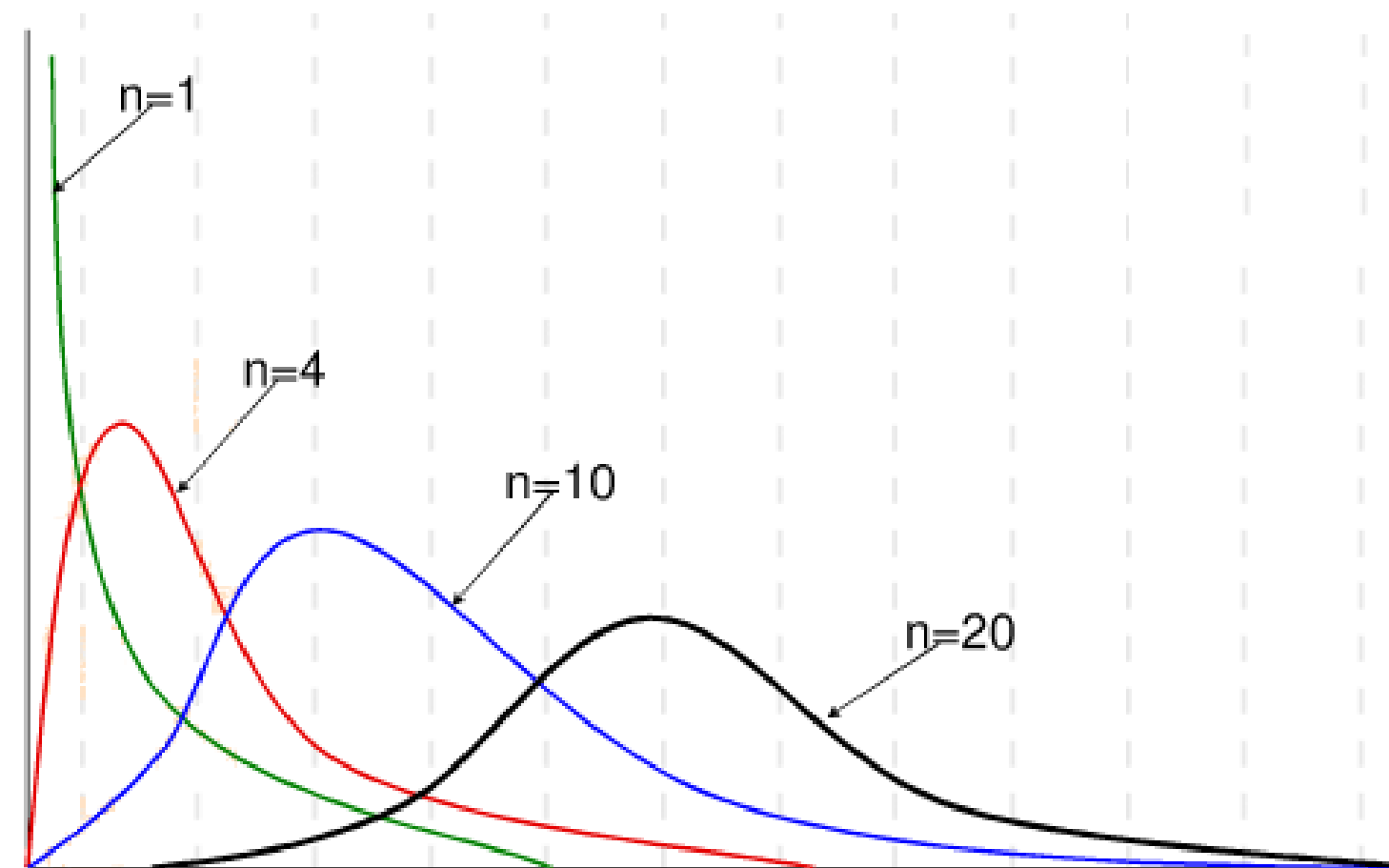
$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

数学期望:

$$E(\chi^2) = n$$

方差:

$$D(\chi^2) = 2n$$



χ^2 分布-实例



电池



抽奖机

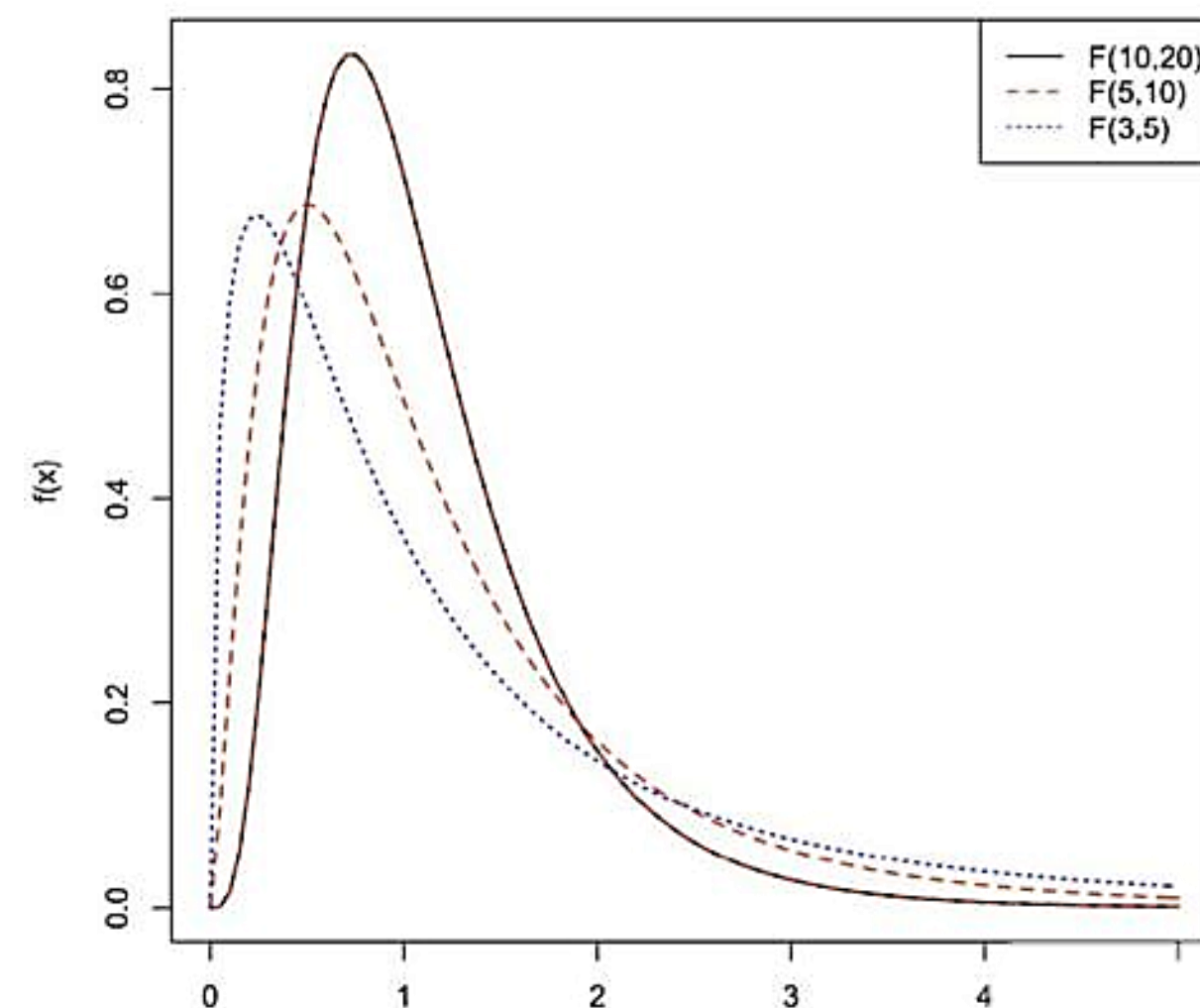
F分布

■ 定义

设 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 且 U, V 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

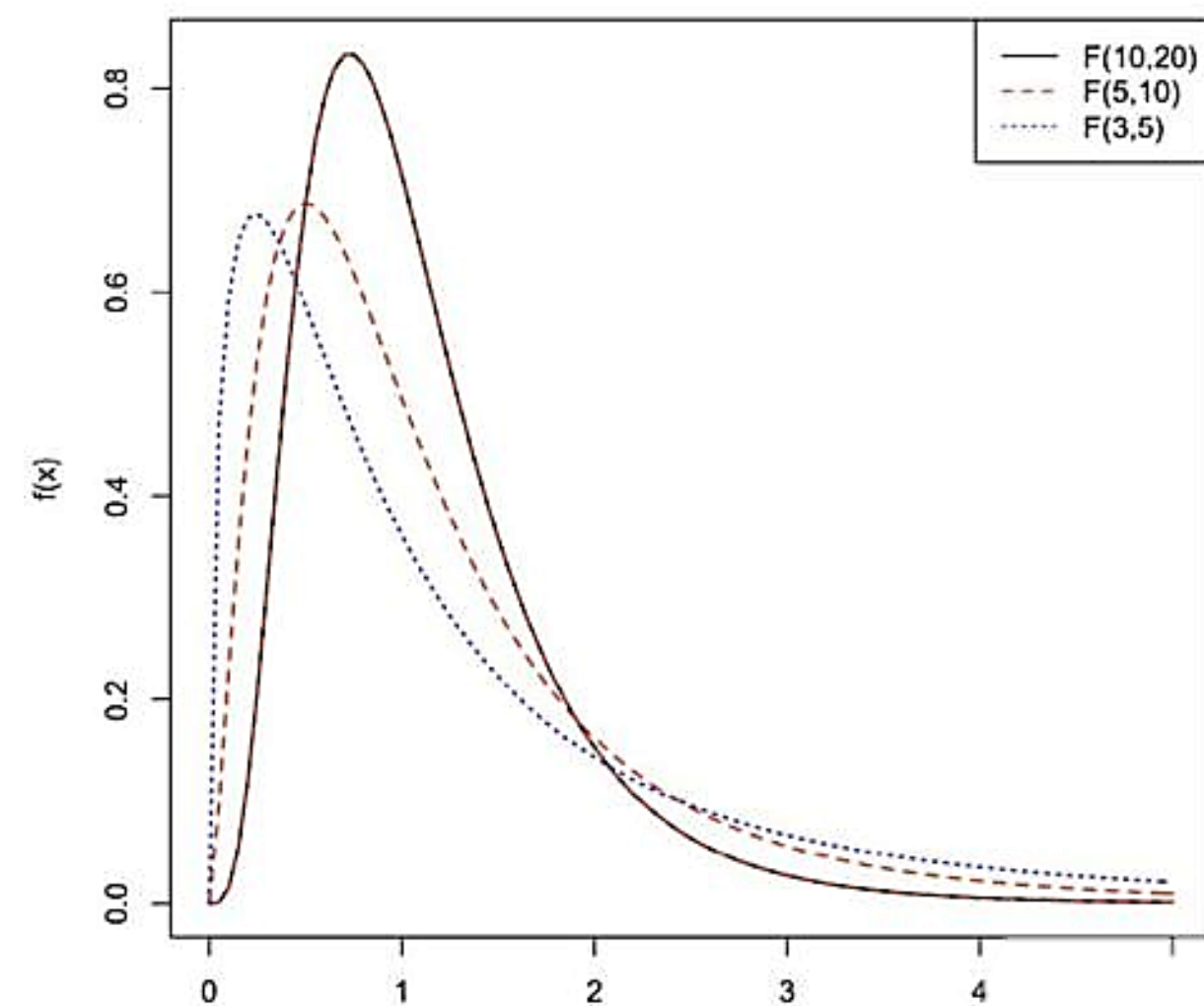
服从自由度为 (n_1, n_2) 的 F 分布, 记为 $F \sim F(n_1, n_2)$



F分布

■ 概率密度

$$\psi(y) = \begin{cases} \frac{\Gamma\left[\frac{(n_1 + n_2)}{2}\right] \left(\frac{n_1}{n_2}\right)^{n_1/2} y^{\left(\frac{n_1}{2}\right)-1}}{\Gamma\left(\frac{n_2}{2}\right) \Gamma\left(\frac{n_1}{2}\right) \left[1 + \left(\frac{n_1 y}{n_2}\right)\right]^{(n_1+n_2)/2}}, & y > 0 \\ 0, & \text{其他} \end{cases}$$

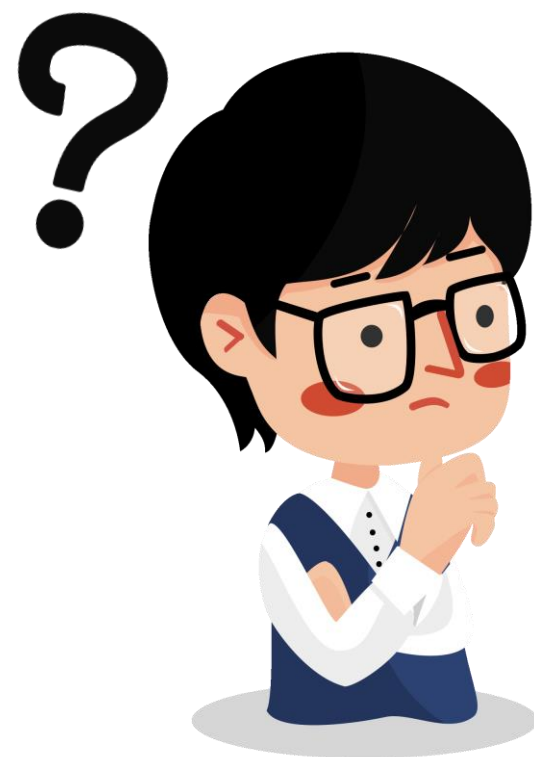


F分布-实例



设两家银行储户的年存款余额均服从正态分布，经市场调查，分别抽取容量为21和16的样本，得样本均值分别为650元和800元，样本方差分别为 80^2 和 70^2 ，能否认为第二家银行储户的平均年存款余额显著高于第一家银行（取 $\alpha = 0.10$ ）？

思考题



01

正态分布的特色都有哪些？

02

哪些场景的数据属于 t 分布？

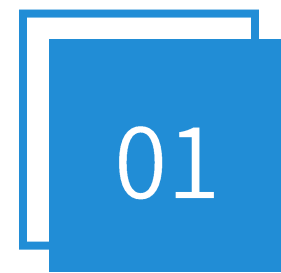
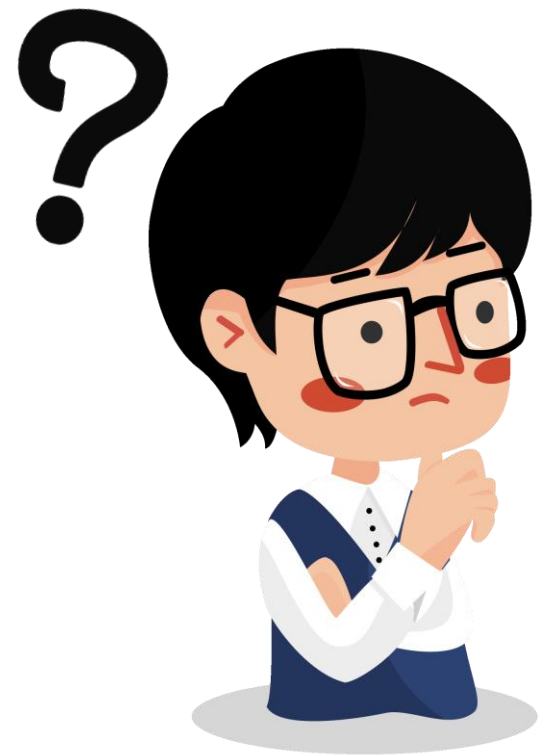
03

t 分布和正态分布的主要区别都有哪些？



总结

思考题



抽样调查遵循的原则是 ()?

- A、随意性原则
- B、可比性原则
- C、随机原则
- D、全面性原则

总结

本章包含两小节内容：

第一节

- 随机样本的定义

第二节

- 抽样分布： t 分布、正态分布、 χ^2 分布、 F 分布

谢谢观看

参考书目：概率论与数理统计·第四版（浙江大学）高等教育出版社