**考点15：统计基本概念（2%）**

**统计学含义及其应用（掌握）**

统计学是一门收集、处理、分析、解释数据并从数据中得出结论的科学

统计学分析数据的方法：描述性统计分析、推断性统计分析

1、描述性分析：研究数据收集、处理和描述的统计学方法，如总体规模、对比关系、集中趋势、离散程度、偏态、峰态……

2、推断性分析：研究如何利用样本数据来推断总体特征的统计学方法，如估计、假设检验、列联分析、方差分析、相关分析、回归分析……

**统计学的基本概念：数据（掌握）**

统计学的对象是数据，数据的形式可以是数字或文字

一、按照计量尺度分类

1、分析型数据：对事物进行分类的结果，如人的性别分为：男、女，中国、美国

特点：定性数据，不可排序，不可计算

2、顺序性数据：对事物类别顺序的测度，如产品分为：一等品、二等品、三等品，优、良、中、差

特点：定性数据，可排序，不可计算

3、数值型数据：对事物的精确测度，如身高为：175cm、180cm

特点：定量数据，可排序，可计算

注1：数据从低级到高级1->2->3，数据可携带信息越多，数据越高级

注2：复数a+bi，不可排序，可计算

注3：区间（分组的数值型数据）属于数值型数据

注4：不同类型数据之间可以转换，低级数据的方法高级数据可以用，高级数据的方法低级数据不可以用

二、数据的其它分类

1、按来源不同：直接来源、间接来源

2、按收集方式不同：观测数据、实验数据

3、按与时间关系不同：截面数据、时间序列数据、混合数据（面板数据）

4、按概型不同：离散型数据、连续型数据

5、一种特殊的数据：虚拟变量数据

**统计学的基本概念：总体、样本、参数、变量（掌握）**

所有和总体有关的都是定值，所有和样本有关的都是变量

1、总体：指研究的所有元素的集合，其中每个元素称为个体

2、样本：从总体中抽取的一部分元素的集合，构成样本的元素的数目称为样本容量

3、参数：指研究者想要了解的总体的某种特征值，主要有：总体均值μ，总体标准差σ，总体比例π

4、统计量：指根据样本数据计算出来的一个量，即样本的某个特征值，主要有：样本均值，样本标准差s，样本比例p

5、变量：指描述事物某种特征的概念，如商品销售额、受教育程度、产品的质量等级

6、变量与数据的关系：变量的具体表现称为变量值，即数据

7、变量的分类：根据变量的数据计量尺度不同来分类

分类变量：类别

顺序变量：有序类别

数值型变量：数字特征

**考点16：数据的描述性统计（3%）**

**应用：能够应用描述性统计知识对业务数据进行恰当的数据特征描述，针对数据描述特征，阐述业务问题、探索问题原因、提出解决问题方法**

思考：某超市后台记录了一年内53万余条消费者的消费数据，请问如何做描述统计分析？（撰写一份数据描述统计分析报告）

五个角度：

**一、总体规模的描述——总量指标**

总量指标：反映在一定时间、空间条件下，某种现象的总体规模、总水平或总成果的统计指标。如：营业额、利润

**二、对比关系的描述——相对指标**

相对指标：两个有相互联系的指标数值之比。如：目标完成率（实际完成/计划完成）

**三、集中趋势的描述——平均指标（了解）**

集中趋势：一组数据向其中心靠拢的趋势。测定集中趋势就是寻找数据水平的代表值或中心值

分类型数据：众数

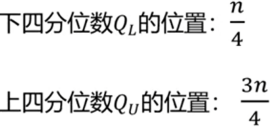
顺序性数据：众数、分位数

数值型数据：众数、分位数、均值

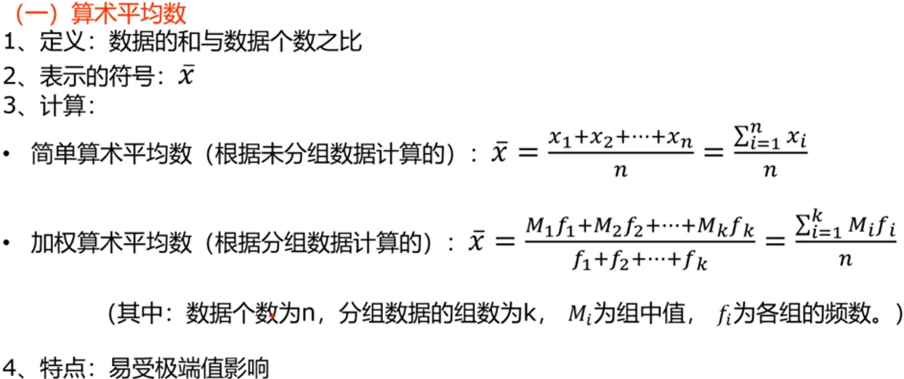
1、众数**（了解）**：出现次数最多的变量值，表示符号M0 计算方式：寻找数据中出现次数最多的值

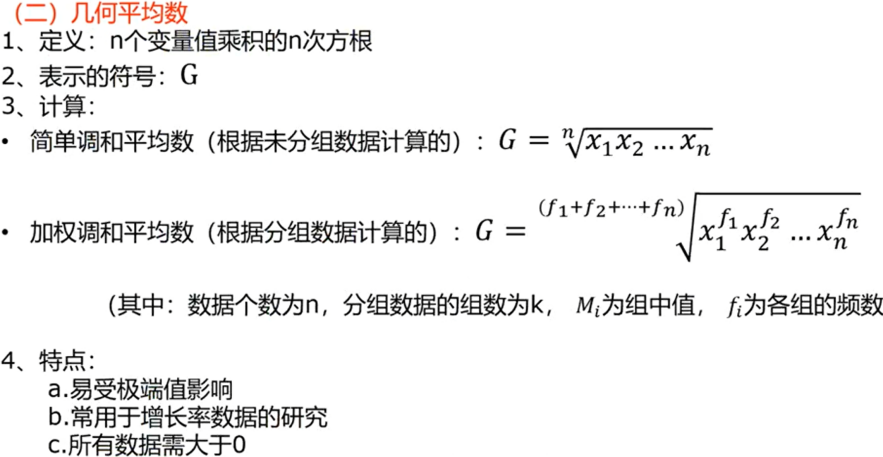
2、分位数**（了解）**：根据数据位置进行划分，处于某些特定位置上的数，常用的有：**二分位数（中位数）、四分位数**、十分位数、百分位数等

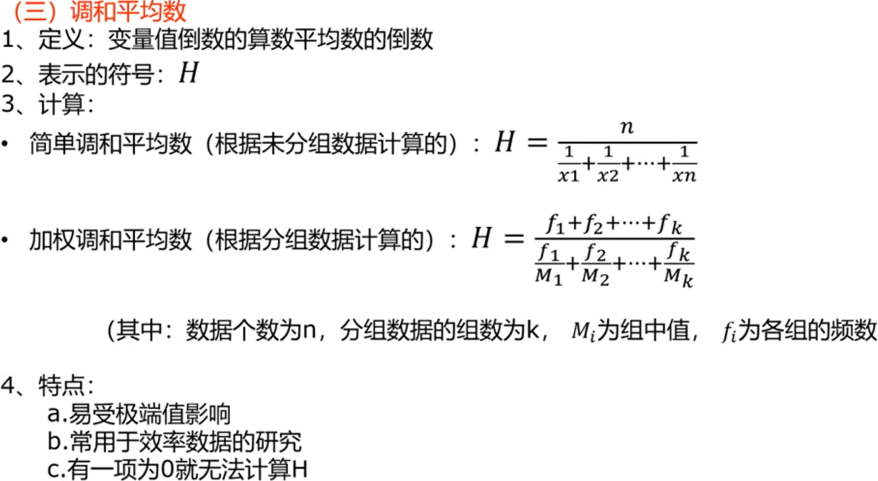
①中位数（二分位数）：数据排序后，处于中间位置上的值，表示符号Me 计算方式：，其中n是数据个数

②四分位数：数据排序后，下四分位数QL是处于25%位置的值，上四分位数QU是处于75%位置的值

3、均值**（了解）**







均值不等式：对于同一组数据，一定满足：算数平均数≥几何平均数≥调和平均数，当且仅当所有数据取相同值时，等号成立

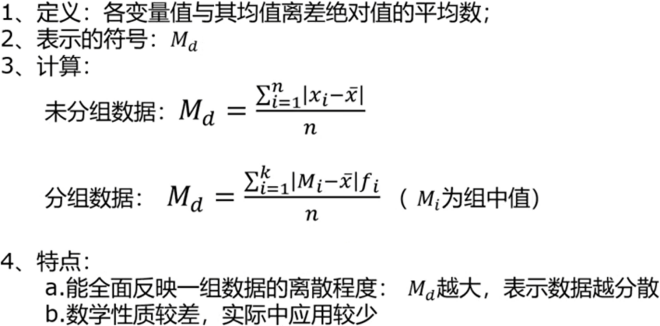
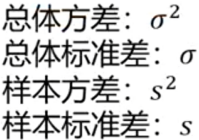
**四、离散程度的描述——变异指标（了解）**

离散程度：反映各变量值远离其中心值的程度，侧面反映了集中趋势的代表程度

1、极差**（了解）**：一组数据的最大值与最小值之差，极差R=max(xi)-min(xi)

特点：①离散程度的最简单测度值；②极易受极端值影响；③未考虑数据的分布

2、平均差（离差）

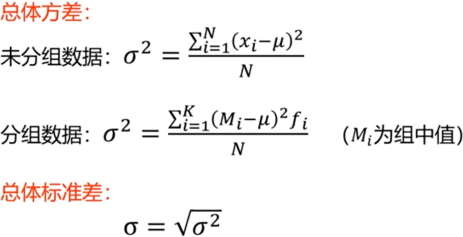
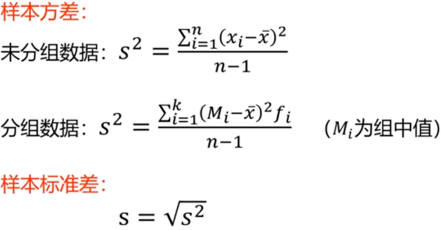


3、方差和标准差**（了解）**

根据总体数据计算的，称为总体方差、总体标准差；根据样本数据计算的，称为样本方差、样本标准差

定义：变量值与其算数平均数的离差的平方的算数平均数

特点：①数据离散程度的最常用测度值；②反映了各变量值与均值的平均差异：方差或标准差越大，表示变量值与均值的平均差异越大

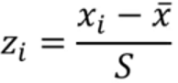
 

总样本数n，样本方差自由度（n-1）：一组数据中可以自由取值的数据的个数

4、离散系数（变异系数）**（了解）**

定义：标准差与均值之比，离散系数

①是对数据相对离散程度的测度；②消除了数据水平不同和数据计量单位不同对数据离散程度的影响；③常用于对不同组别数据离散程度的比较

5、标准化值（标准分数）：对某一数据在全体中相对位置的度量，可用于判断一组数据是否有离群值，用于对变量的标准化处理

①对于对称分布的数据，经验法则：

约有68%的数据在平均数加减1个标准差的范围之内

约有95%的数据在平均数加减2个标准差的范围之内

约有99%的数据在平均数加减3个标准差的范围之内

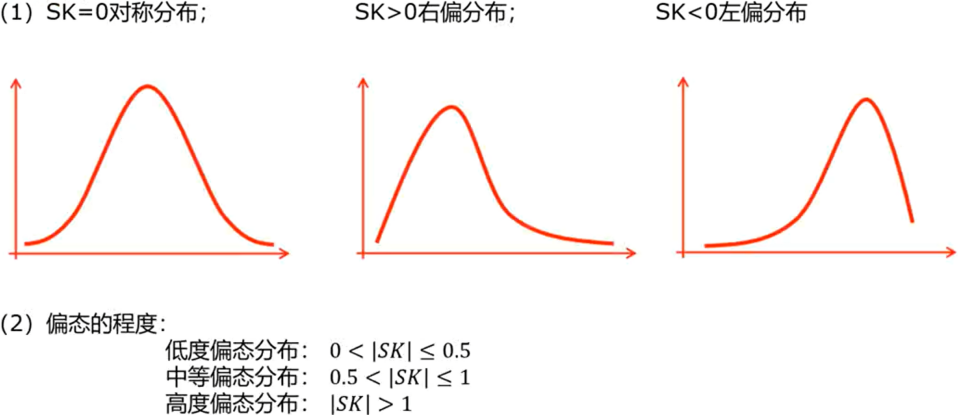
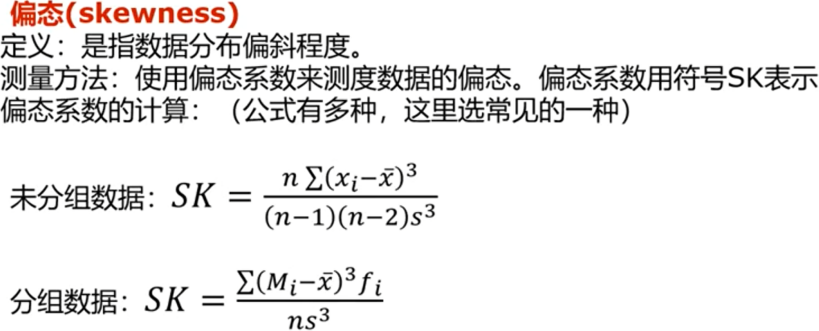
①对于任意分布的数据，切比雪夫不等式：

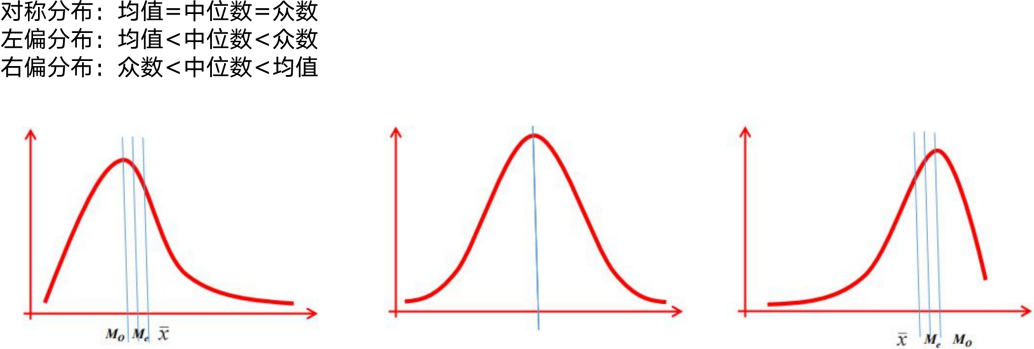
至少有1-1/k2的数据落在平均数加减k个标准差之内，其中k是大于1的任意值，但不一定是整数

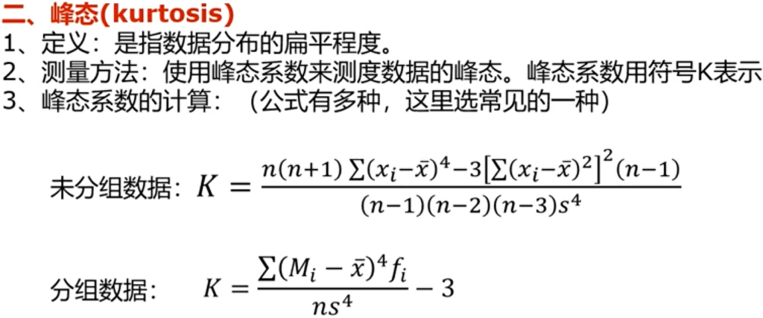
例如，对于k=2，3，4，该不等式的含义是：至少有75%的数据落在平均数加减2个标准差的范围之内；至少有89%的数据落在平均数加减3个标准差的范围之内；至少有94%的数据落在平均数加减4个标准差的范围之内

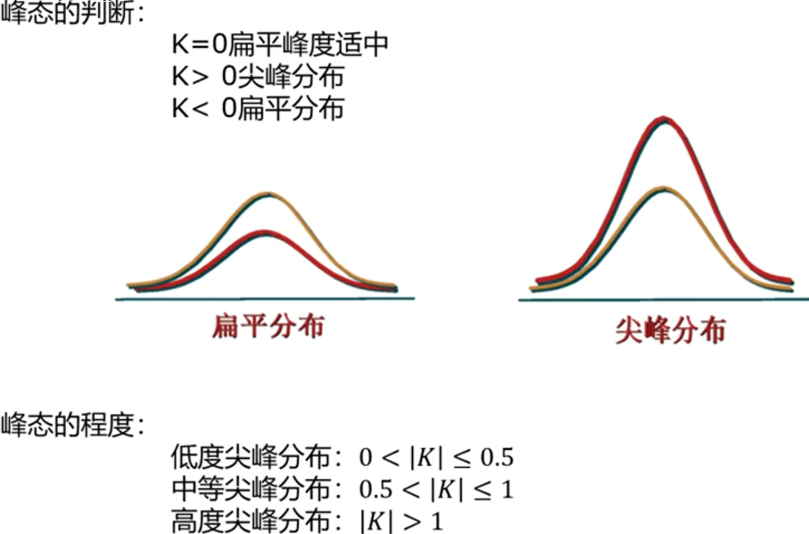
通常选取平均数加减3个标准差之外的数据作为异常值，因为概率只有小概率10%

**五、分布形态的描述——偏态与峰态（了解）**









**六、描述性统计图表（了解）**

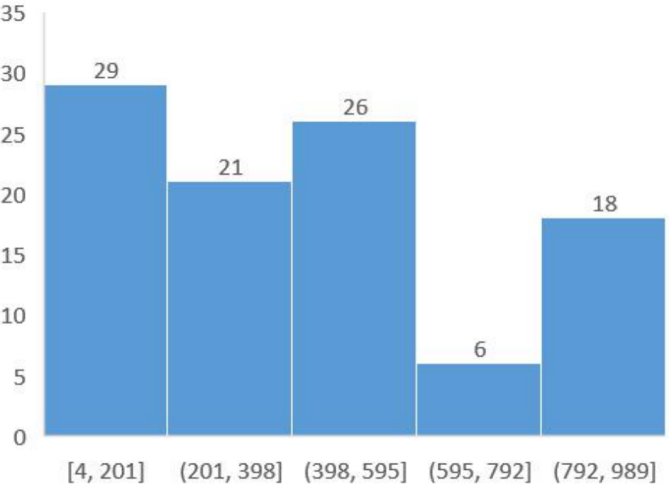
1、直方图：由一系列高度不等的矩形表示数据分布的情况

频数分布直方图：在统计数据时，横轴按组距分类，纵轴表示频数，每个矩形的高代表对应组距数据的频数

组数：把数据按照不同的范围分成几个组，分成的组的个数称为组数

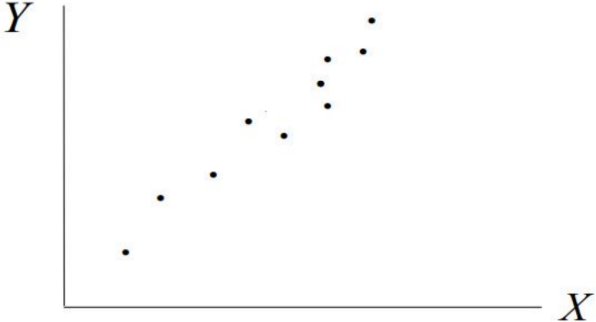
组距：每一组数据的极差

特点：能够显示各组频数分布的情况，易于显示各组之间频数的差别

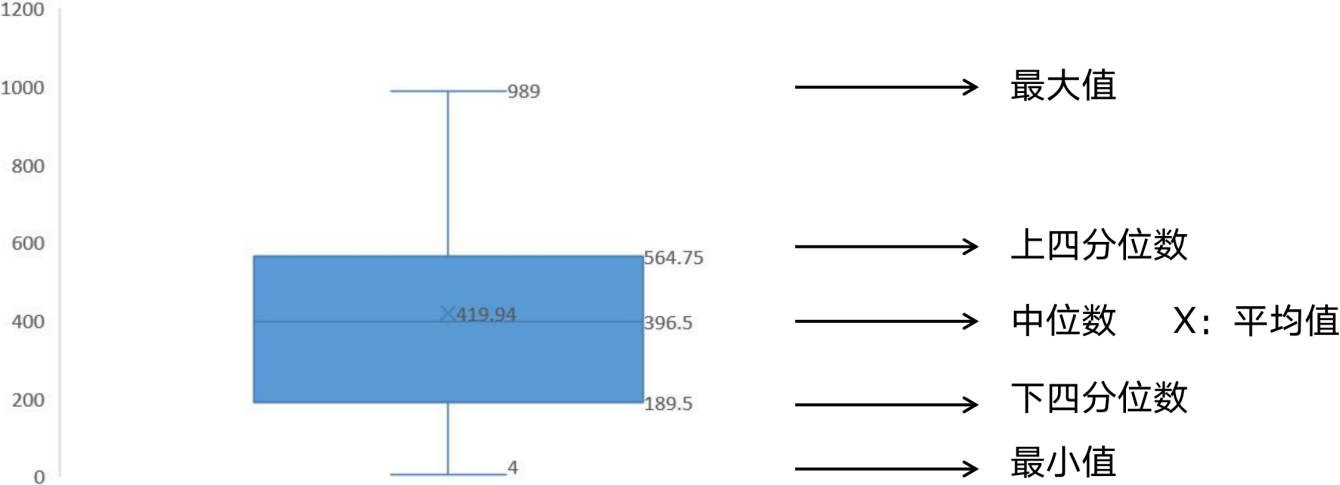


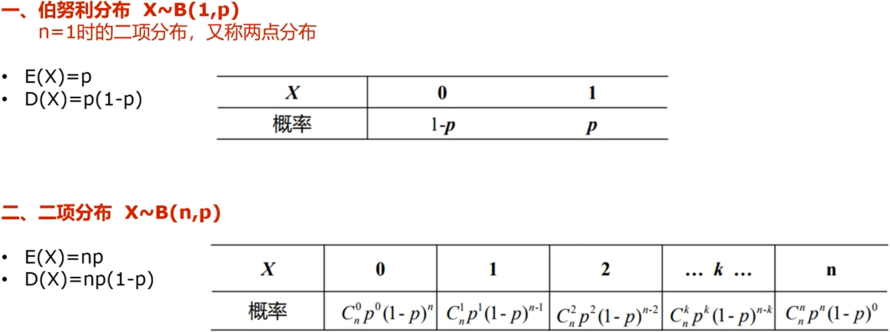
2、散点图：数据统计分析中，数据点在平面直角坐标系上的分布图，表示因变量随自变而变化的大致趋势

特点：展示数据的分布情况，发现变量之间的关系



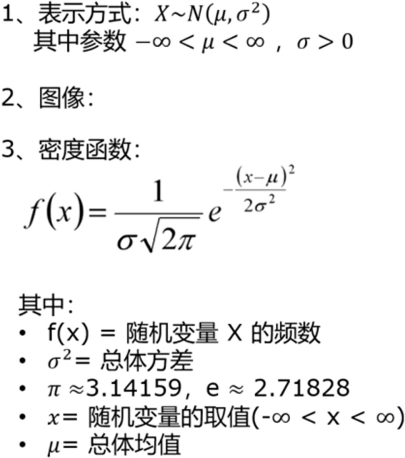
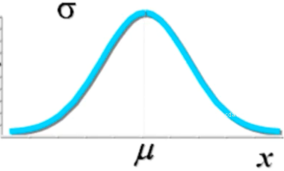
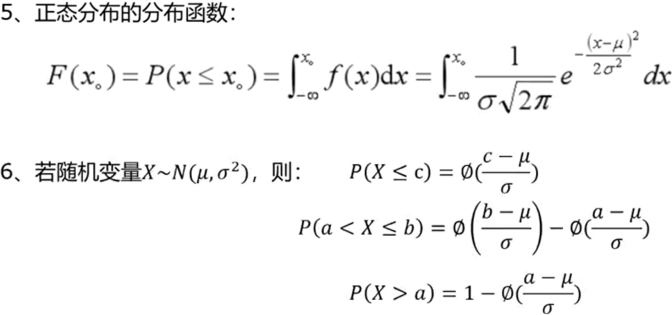
3、箱型图（盒须图/箱线图）：显示一组数据分散情况的统计图

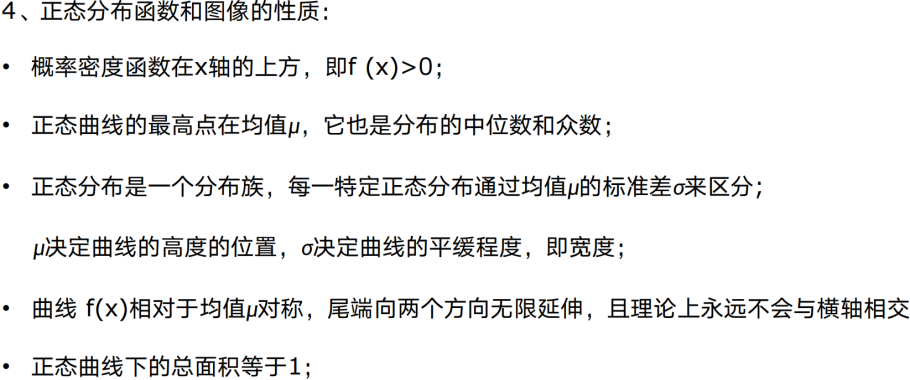


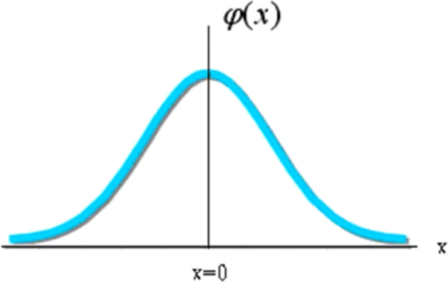
**考点17：统计分布（3%）**

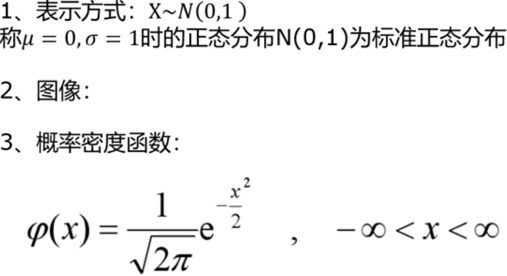
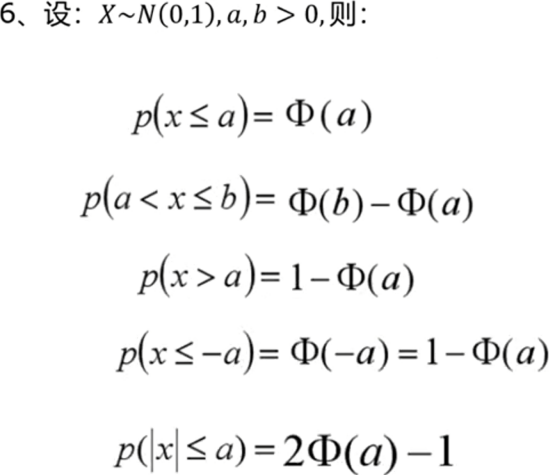
**离散分布：两点分布，二项分布（掌握）**

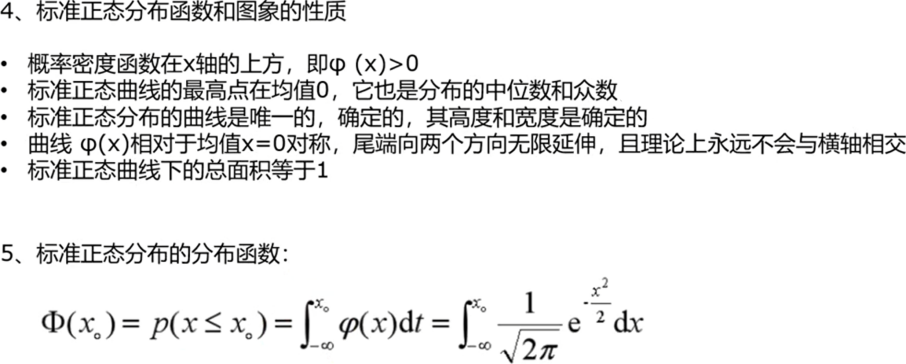
**连续分布：正态分布（掌握）**

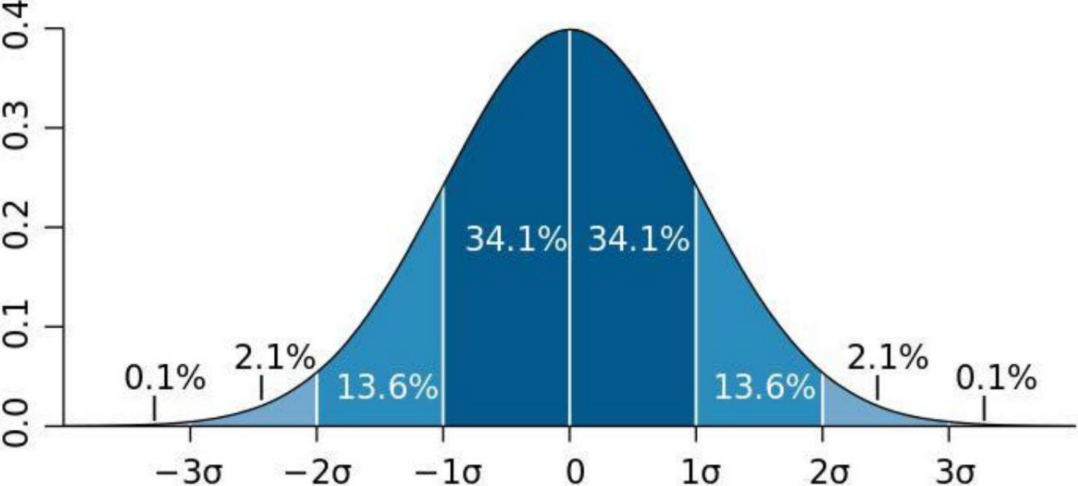


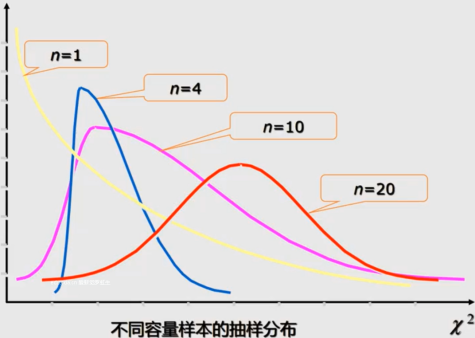


**连续分布：标准正态分布（掌握）**





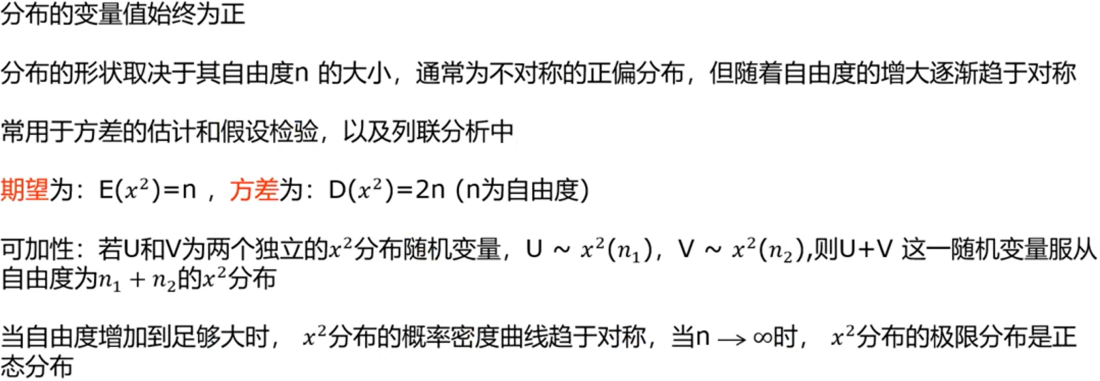


**抽样分布：X2分布**（多个标准正态分布的平方和）**（掌握）**

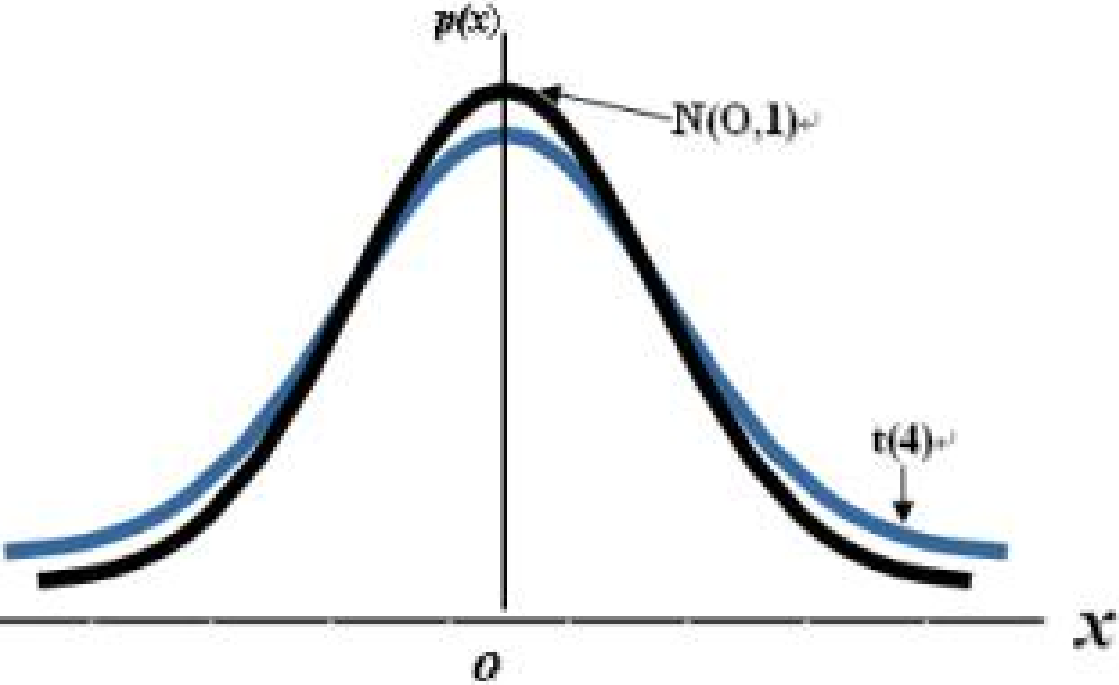
设X ~ N（μ，σ2），则

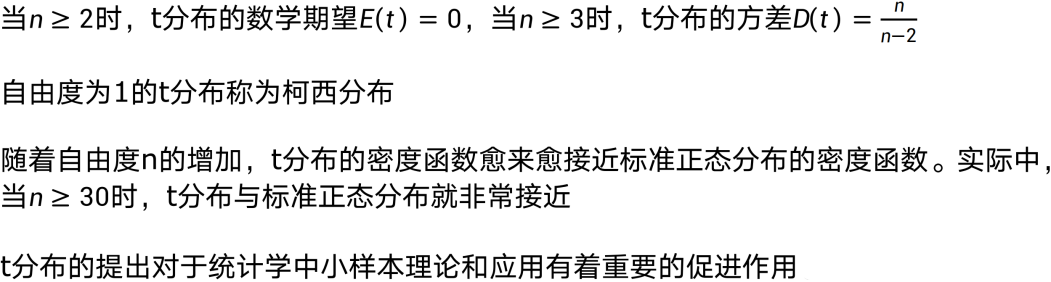
令Y = Z2，则Y服从自由度为1的X2分布，即Y ~ X2（1）

当总体X ~ N（μ，σ2），从中抽取容量为n的样本，则样本方差

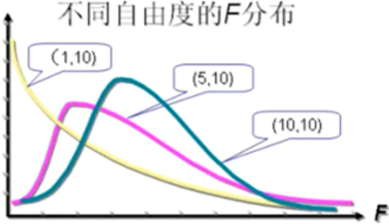


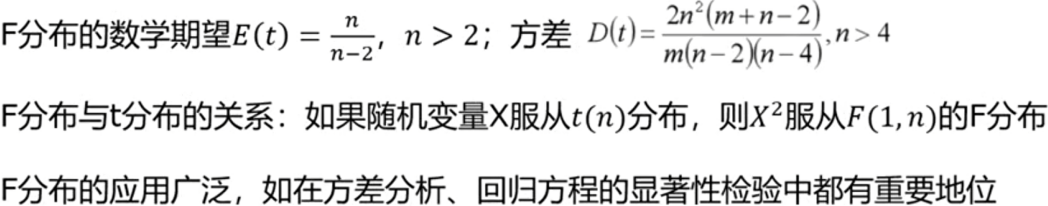
**抽样分布：t分布（掌握）**

设随机变量X ~ N（0，1），Y ~ X2（n），且X与Y相互独立，则称为自由度为n的t分布，记为 t (n)



**抽样分布：F分布（掌握）**

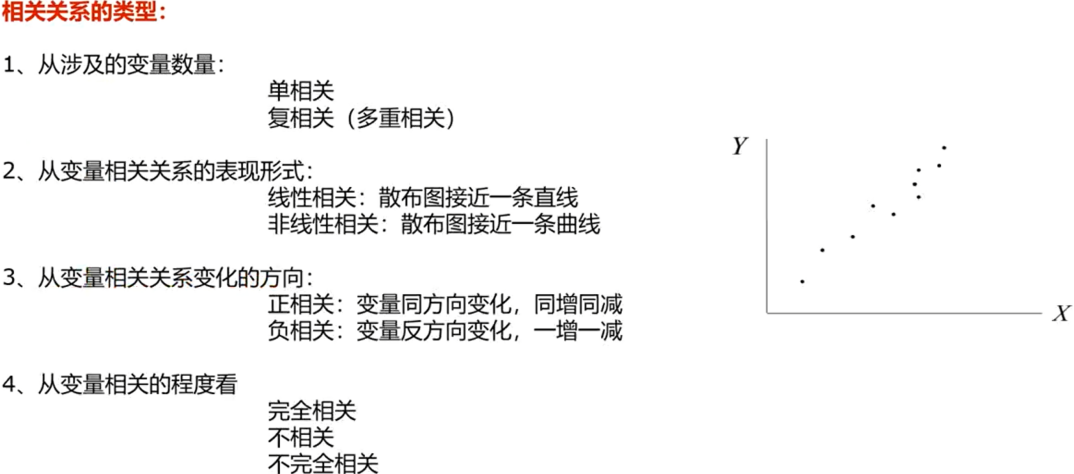
若U ~X2（m），V ~X2（n）且U与V相互独立，则为服从自由度m和n的F分布，记F~F（m，n），m分子自由度/n分母自由度



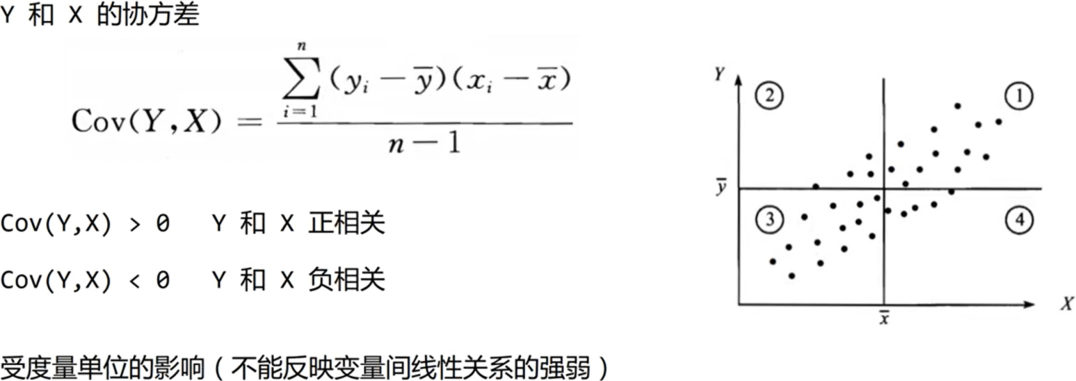
**考点18：相关分析（2%） **

**相关关系的描述——散点图（掌握）**

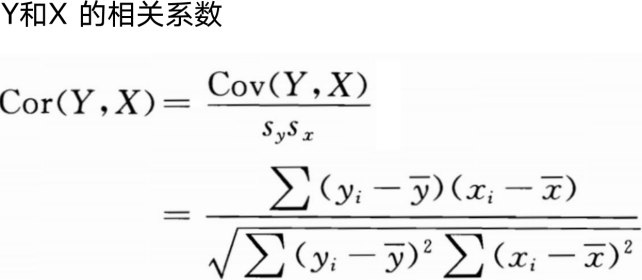
**相关关系的类型（掌握）**

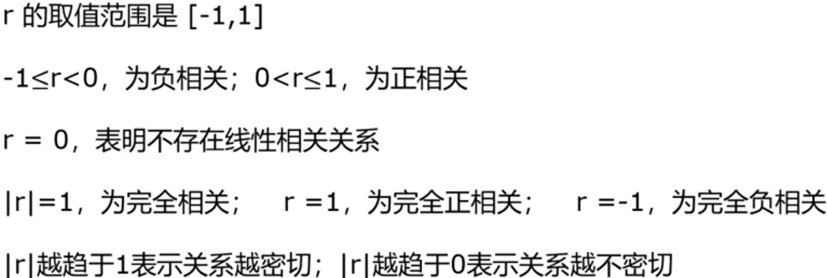


**相关关系的度量——协方差Cov （掌握）**



**相关关系的度量——相关系数Cor**（只能度量线性相关性）**（掌握）**





相关系数易受到离群值的影响，实际应用中需要先去除离群值再进行相关分析