§5.2 总体与样本

一 总体和个体

总体——研究对象的全体元素组成的集合. 通常它们是一些数量指标，其数值随着个体的不同而不同，例如：研究某城市在职职工的年收入情况，研究北京市中学生的身高与体重情况.

在进行理论研究时，我们将研究的数量指标视为随机变量X.（或随机向量$X=(X_1,\cdots,X_n)$).

个体——组成总体的每一个元素. 即总体的每个数量指标，用X_i表示.

总体可分为有限总体和无限总体.

二　样本

样本---从总体中抽取的待检个体的集合.

样本容量---样本中所包含的个体数目.

容量为n的样本可以看作n维随机变量. 但是, 一旦取定一组样本，得到的是n个具体的数$(x_1, x_2, \cdots, x_n)$，称为样本的一次观察值，简称样本值.

样本空间---样本所有可能取值的集合.

由于抽样的目的是为了对总体进行统计推断，为了使抽取的样本能很好地反映总体的信息，必须考虑抽样方法.

如何科学地进行抽样检查?

一个著名的案例

在抽样调查中，样本的选择是至关重要的，样本能否代表总体，直接影响着统计结果的可靠性。下面的故事是一次著名的失败的统计调查，被称为抽样中的泰坦尼克事件。它可以帮助我们理解为什么一个好的样本如此重要。

在1936年美国总统选举前，一份颇有名气的杂志的工作人员做了一次民意调查。调查兰顿（当时任堪萨斯州州长）和罗斯福（当时的总统）中谁将当选下一届总统。为了了解公众意向，调查者通过电话簿和车辆登记簿上的名单给一大批人发了调查表（注意在１９３６年电话和汽车只有少数富人拥有）。通过分析收回的调查表，显示兰顿非常受欢迎，于是杂志预测兰顿将在选举中获胜。

实际上选举结果正好相反，最后罗斯福在选举中获胜，其数据如下：

候选人	预测结果	选举结果
罗斯福	43％	62％
兰顿	57％	38％

最常用的一种抽样方法叫作“简单随机抽样”，它要求抽取的样本满足下面两点：

1. 代表性：$X_1, X_2, \cdots, X_n$中的每个X_i与所考察的总体有相同的分布.

2. 独立性：$X_1, X_2, \cdots, X_n$是相互独立的.

若不特别说明，本文都是指简单随机样本.

具体方法：

1、抽签法　编号、标签、均匀搅拌、抽签。

2、随机数表法

3、分层抽样等

个体

待检查的个体

样本

$$(X_1, X_2, \cdots, X_n)$$

$$(x_1, x_2, \cdots, x_n)$$

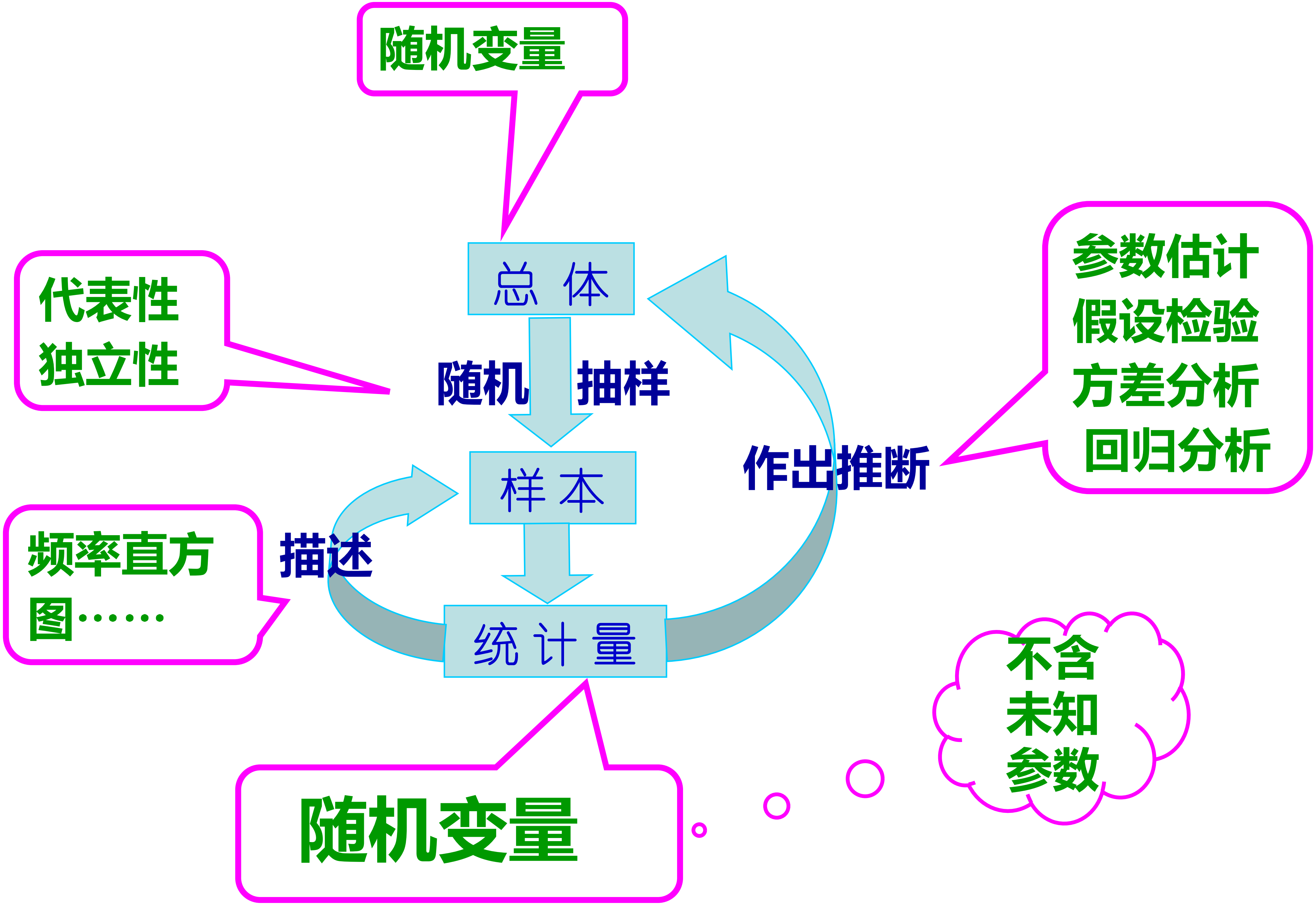
三　样本的分布

设总体X的分布函数为$F(x)$，则样本$(X_1,\cdots,X_n)$的联合分布函数为

$$F_{总}(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$$

若总体X的密度函数为$f(x)$，则样本$(X_1,\cdots,X_n)$的联合密度函数为

$$f_{总}(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i).$$


随机变量
代表性
独立性
总 体
随机 抽样
样 本
参数估计
假设检验
方差分析
回归分析
作出推断
频率直方图……
描述
统 计 量
不含未知参数
随机变量