
第三章 抽样误差

Sampling Error

主要内容

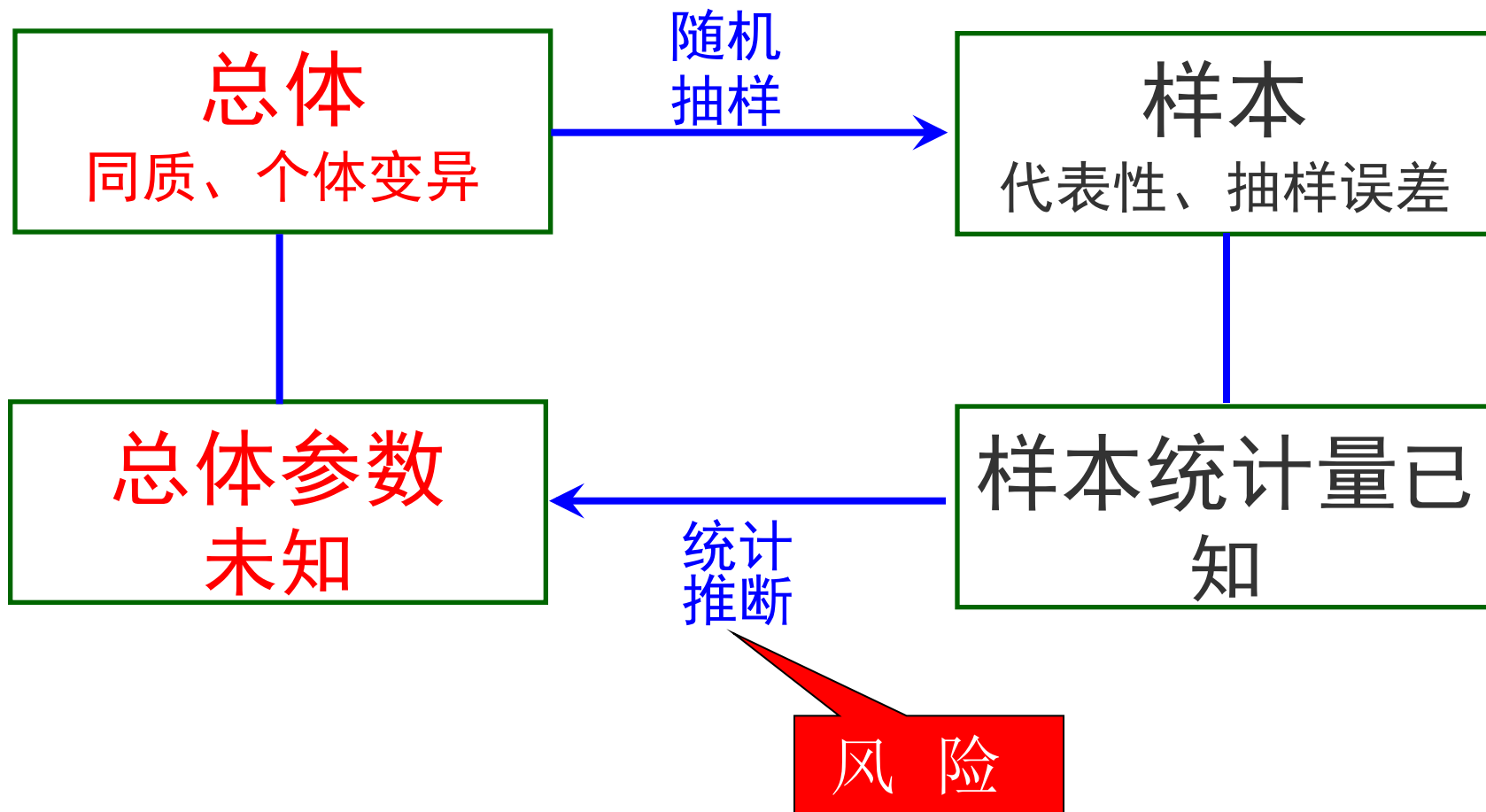
- 抽样误差
 - 中心极限定理
 - 标准误
 - t 分布
 - χ^2 分布
 - F 分布
-

1. 抽样误差

Sampling Error

- 抽样误差
- 中心极限定理
- 标准误
- 统计分布

了解抽样误差的重要性



抽样误差

- **sampling error, sampling variability**
 - 由抽样引起的样本统计量与总体参数间的差别。
 - 原因：个体变异 + 抽样
 - 表现：
 - 样本统计量与总体参数间的差别
 - 不同样本统计量间的差别
 - 抽样误差是不可避免的！
 - 抽样误差是有规律的！

均数的模拟试验

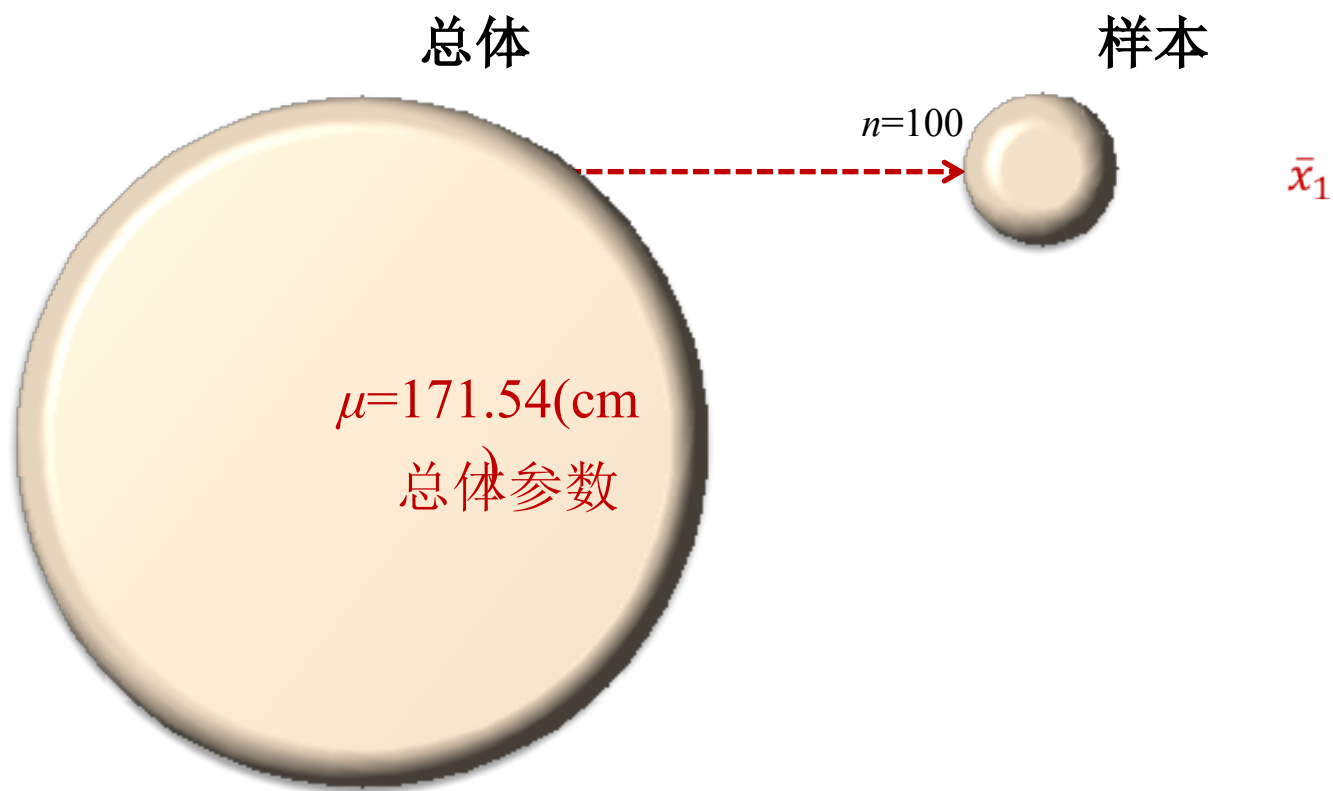
- 假设一个已知总体，从该总体中抽样，对每个样本计算样本统计量(均数、方差等)，观察样本统计量的分布规律——**抽样分布规律**。
 - 正态分布总体
 - 偏三角分布总体
 - 均匀分布总体
 - 指数 F 分布总体
 - 双峰分布总体

均数的模拟试验

■ 考察:

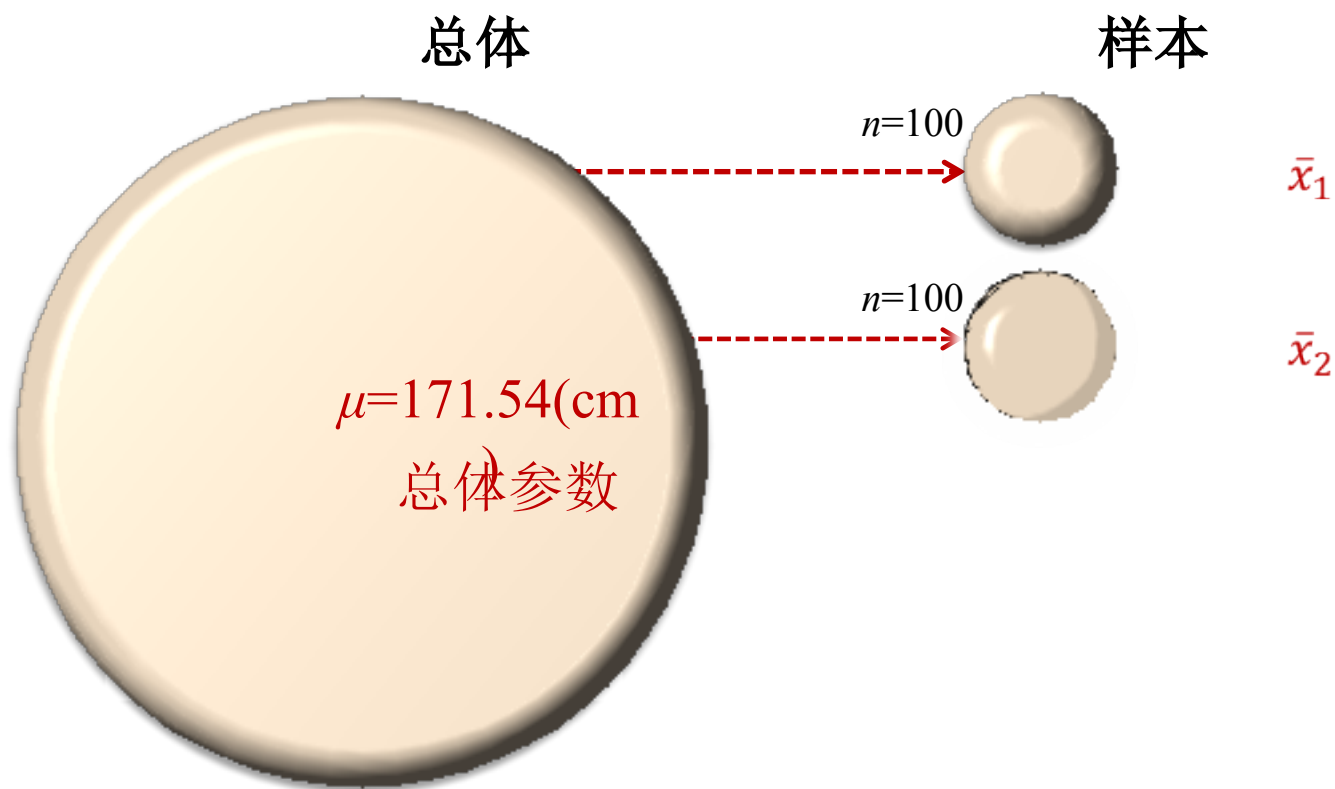
- 样本均数的均数与总体均数有何关系?
- 样本均数的标准差与总体标准差有何关系?
- 样本均数的分布形状如何?
- 不同的样本含量对上述性质的影响如何?

抽样分布规律



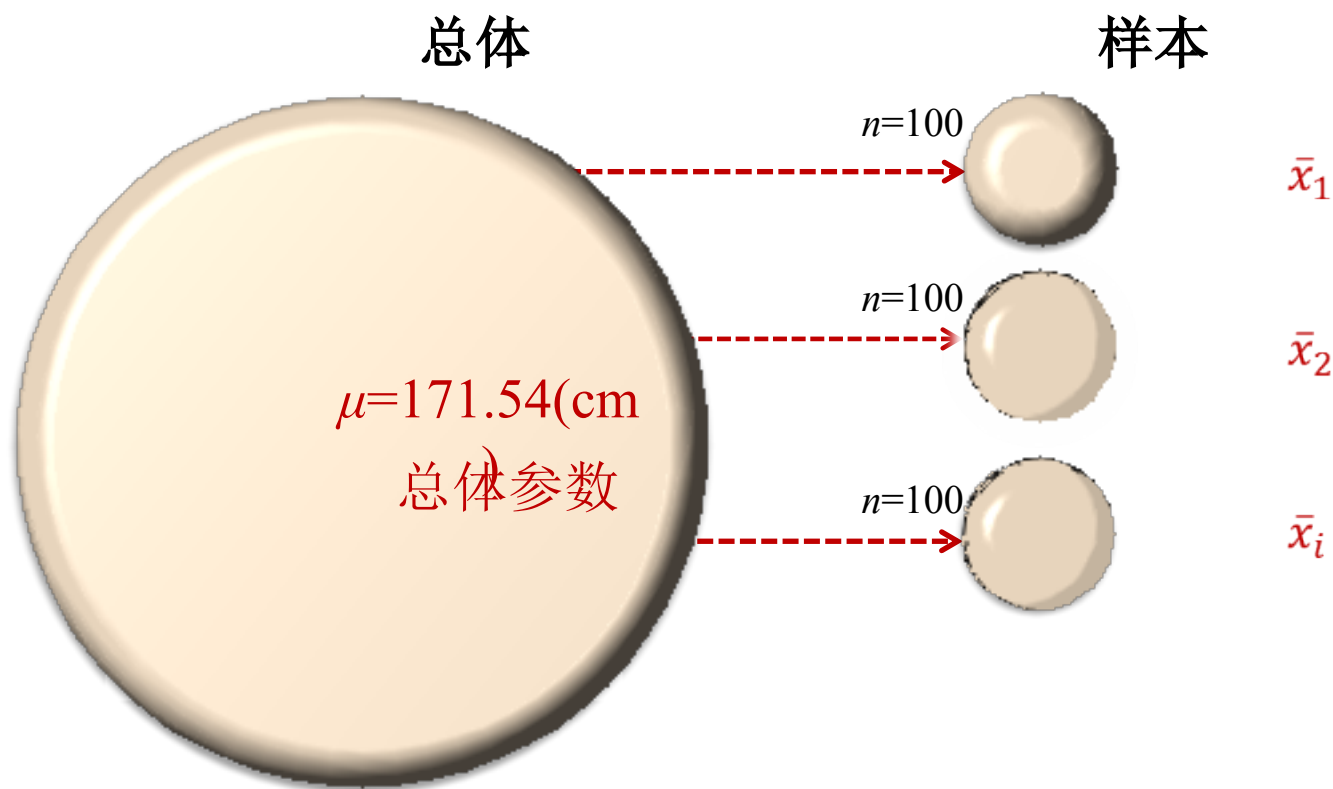
某地健康成年男子身高的模拟抽样研究

抽样分布规律



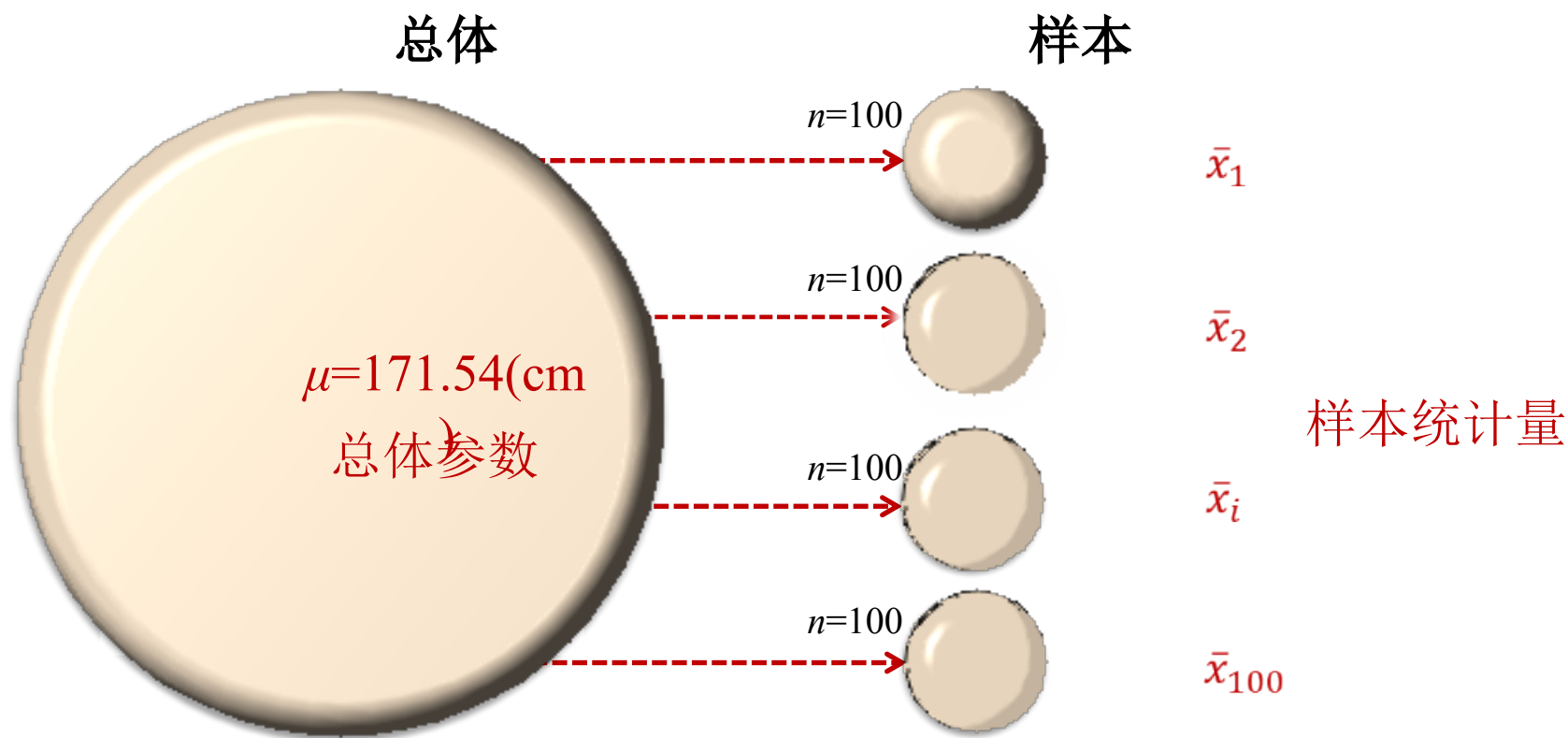
某地健康成年男子身高的模拟抽样研究

抽样分布规律



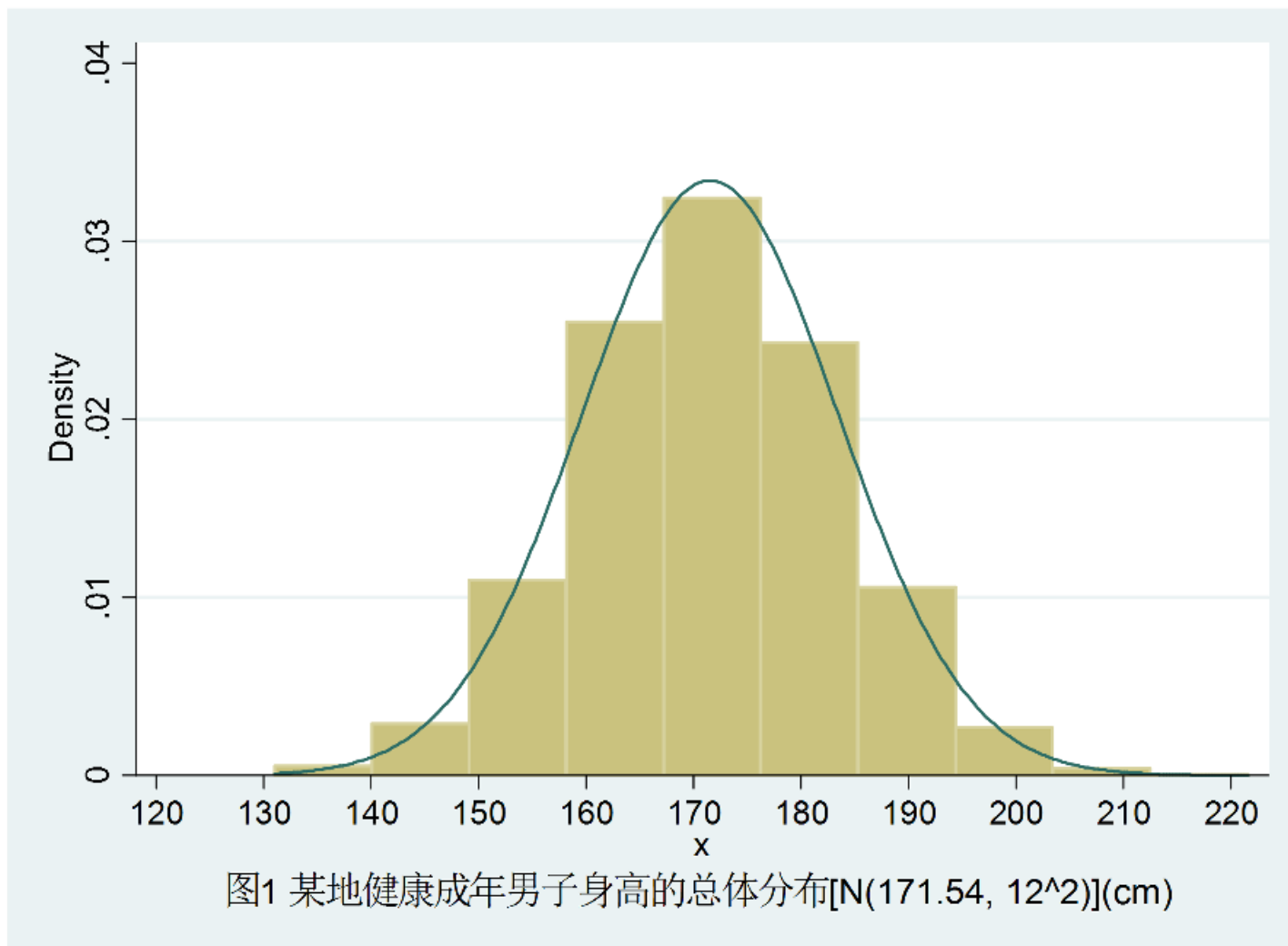
某地健康成年男子身高的模拟抽样研究

抽样分布规律



某地健康成年男子身高的模拟抽样研究

抽样分布规律



抽样分布规律

表 21-1-1 某地健康成年男子身高总体 $N(171.54, 12^2)$ 中模拟抽样的 100 个样本均数结果 (cm)

样本号	均数	标准差	样本号	均数	标准差	样本号	均数	标准差	样本号	均数	标准差	样本号	均数	标准差
1	171.1768	12.5991	21	172.1823	11.6294	41	172.8333	10.15677	61	170.5796	12.1612	81	170.8431	12.9470
2	170.8568	11.1914	22	172.1389	11.4059	42	172.8637	10.98271	62	171.6205	11.5676	82	171.6137	11.8899
3	173.7304	12.1143	23	172.8839	13.0253	43	171.8227	12.09752	63	171.0124	12.2322	83	171.6292	12.2531
4	171.2272	12.7321	24	172.0293	10.2737	44	173.1659	11.33918	64	170.3443	11.2626	84	172.2000	11.9178
5	171.0280	12.7819	25	169.6583	11.1005	45	172.1146	13.08528	65	172.7612	11.4710	85	170.4517	13.6784
6	170.9146	11.5469	26	171.7281	10.9716	46	170.7761	12.38188	66	172.1640	12.8330	86	171.4039	11.7507
7	171.7900	12.4601	27	171.9847	11.1943	47	171.6470	13.01202	67	172.7316	12.8509	87	169.9395	11.5555
8	171.6660	11.6199	28	172.0266	10.8602	48	172.3256	12.48994	68	169.9331	12.4960	88	174.5053	12.4396
9	173.9558	11.3470	29	170.3860	12.5757	49	172.9834	11.13018	69	172.8194	12.5225	89	172.9512	11.1365
10	170.7803	11.5802	30	171.6558	10.8805	50	169.6766	13.02868	70	170.4574	13.0405	90	172.0797	9.3846
11	173.3098	12.9930	31	171.1461	11.7635	51	170.8512	11.34827	71	170.7037	11.6541	91	170.5258	12.8312
12	169.6244	11.4035	32	170.9917	12.1074	52	168.9528	10.67901	72	173.5195	11.9712	92	172.7319	11.2607
13	172.3144	11.7889	33	169.5773	11.8128	53	172.0389	13.14405	73	169.7899	11.0616	93	169.4460	12.4285
14	172.0538	11.3126	34	171.3332	12.0102	54	170.8682	12.90144	74	170.8843	11.4036	94	170.8573	12.1185
15	169.9233	12.0923	35	170.5023	12.9004	55	171.6075	12.36632	75	170.9541	11.6814	95	170.2771	10.9356
16	171.2382	12.1154	36	171.1187	12.8147	56	170.9159	12.1840	76	174.2778	11.3627	96	171.2466	11.1092
17	170.8989	12.0248	37	172.0480	12.3733	57	173.9149	11.34987	77	170.7072	12.6575	97	169.1708	11.4507
18	173.6771	12.9343	38	170.5497	14.4524	58	173.3633	10.87521	78	170.9617	12.3957	98	173.6636	12.8297
19	172.3332	11.3445	39	172.4588	12.3688	59	170.2060	12.80662	79	171.1736	12.6035	99	171.3037	12.5800
20	172.7156	11.1711	40	170.5130	12.8782	60	171.4155	11.59017	80	169.7599	12.3865	100	172.1228	11.4663

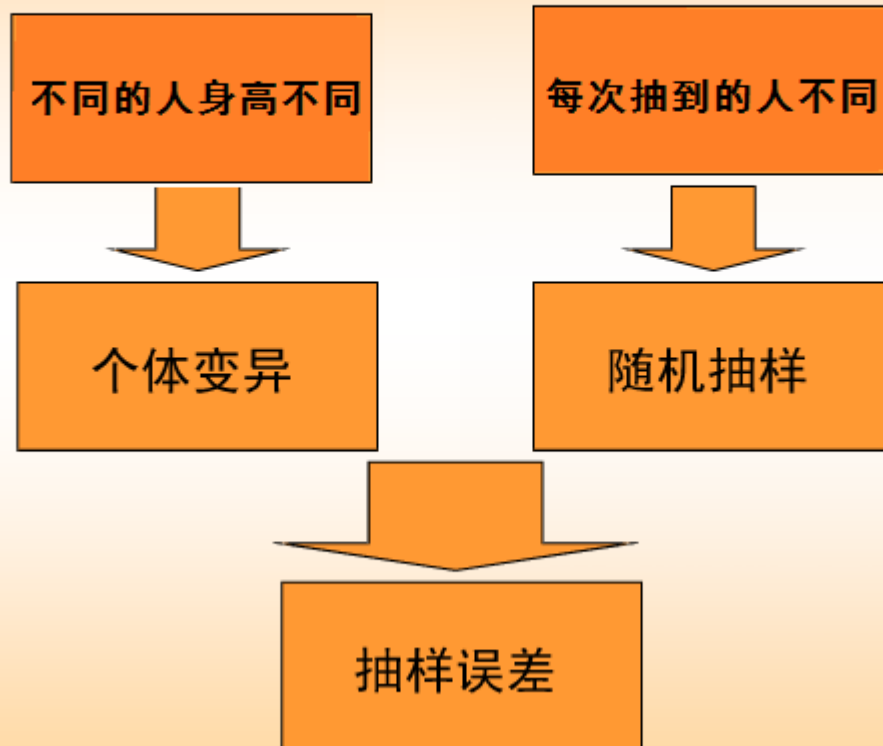
注：样本量 $n=100$ 。

结论 1

- 各样本均数未必等于总体均数；
 - 样本均数间存在差异；
-

抽样分布规律

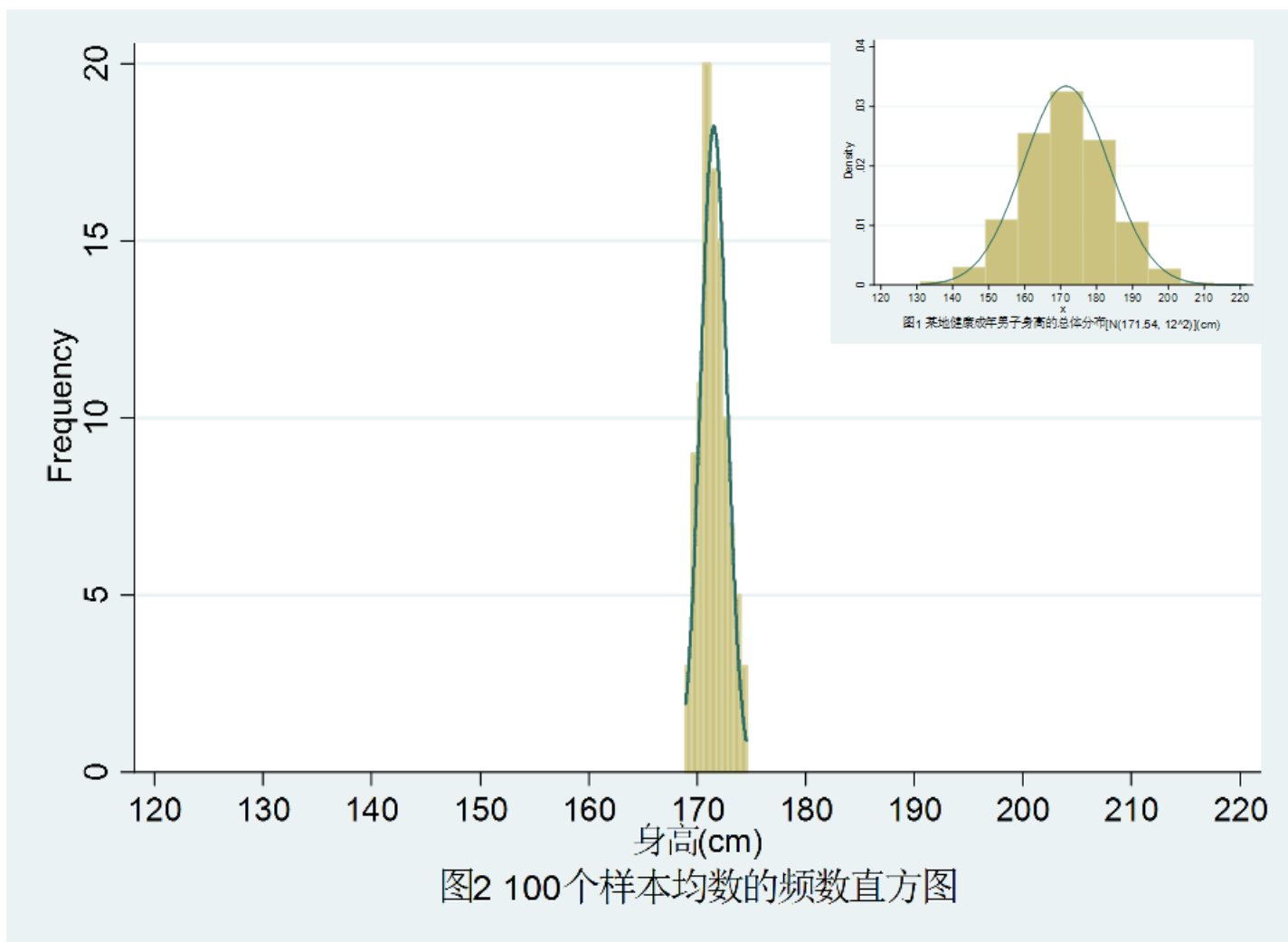
多次抽样得到了不同的结果，原因何在？



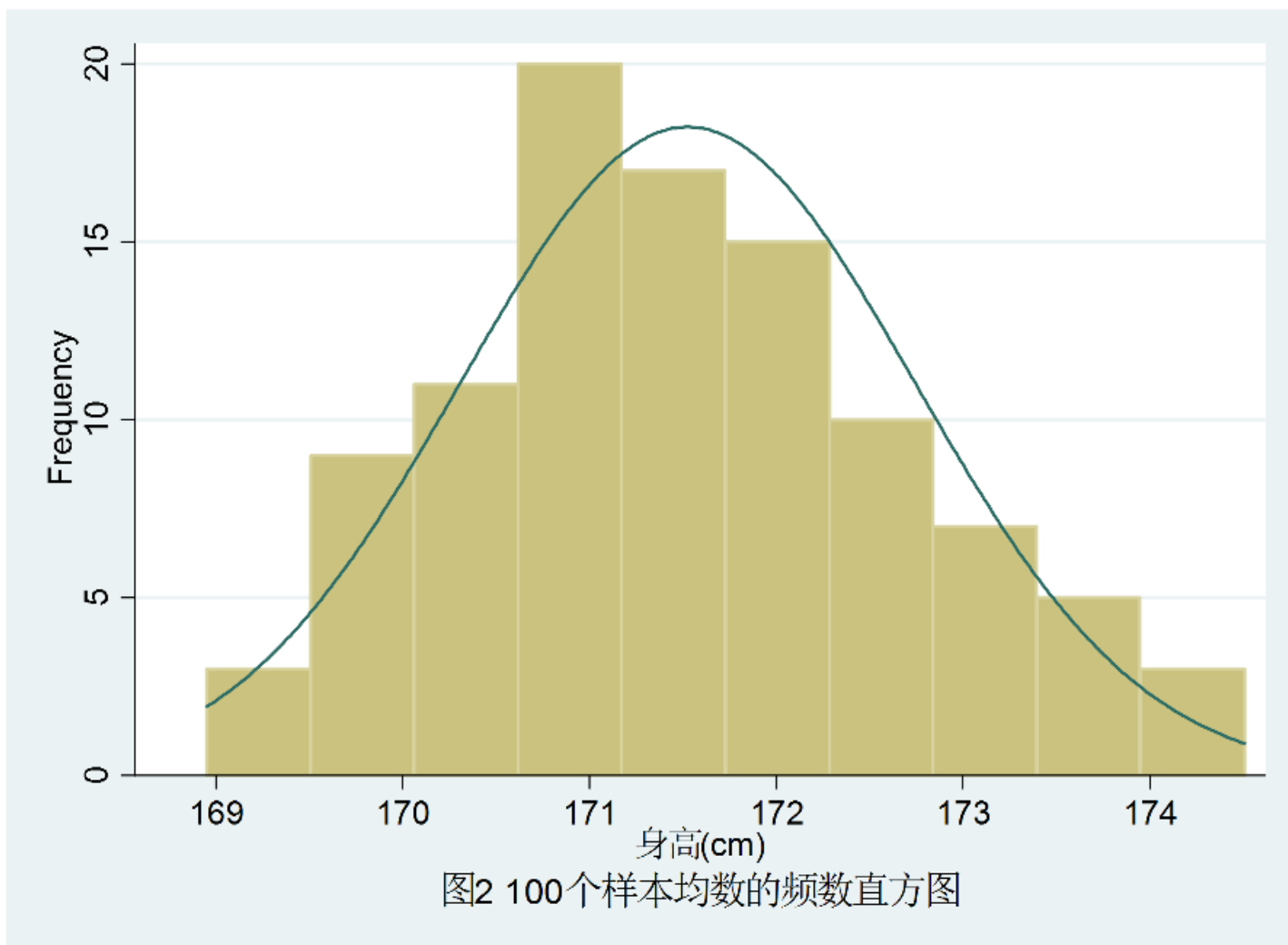
抽样分布规律

- 由抽样实验所得的**100**个样本有何分布特点？
 - 均数分布的频数直方图。
 - 拟合这**100**个数据的分布曲线。

抽样分布规律



抽样分布规律



结论2

- \bar{X} 的分布很有规律，围绕着 μ ，中间多，两边少，左右基本对称；
- 样本均数的变异范围较之原变量的变异范围大大缩小；

2.中心极限定理

central limit theorem

- 抽样误差
- 中心极限定理
- 标准误
- 抽样分布
- 参数估计

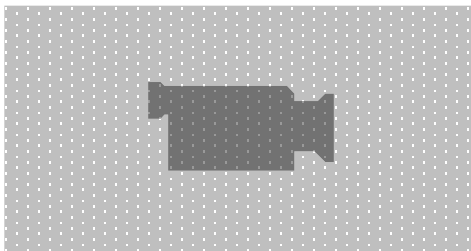
中心极限定理(central limit theorem)

(一) 从均数为 μ 、标准差为 σ 的正态总体中，独立随机抽取例数为 n 的样本，样本均数的分布服从正态分布；

■ 样本均数的均数为 μ ；

■ 样本均数的标准差为 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

中心极限定理





总体分布 A X



$n=2$

X



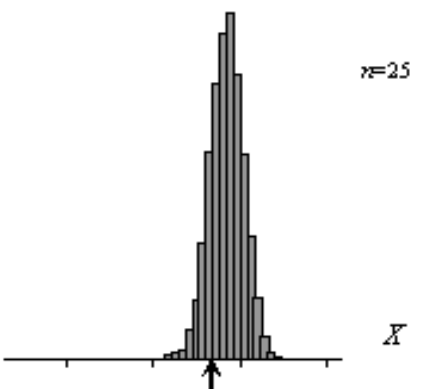
$n=4$

X



$n=10$

X



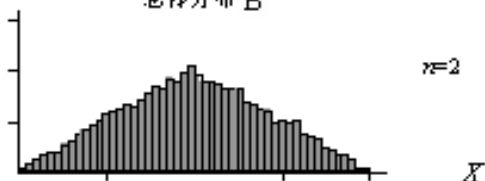
$n=25$

X

\bar{X} 的抽样分布

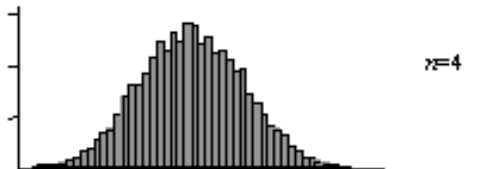


总体分布 B X



$n=2$

X



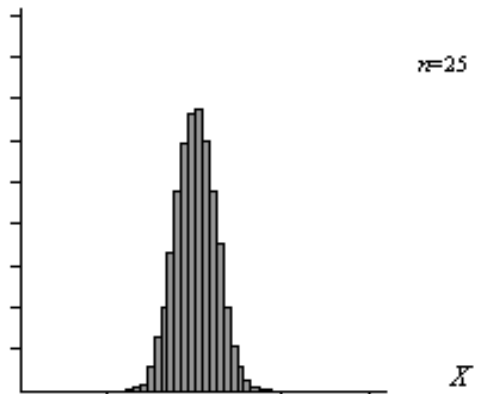
$n=4$

X



$n=10$

X



$n=25$

X

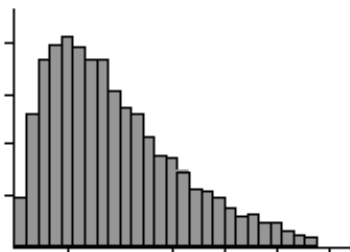
\bar{X} 的抽样分布



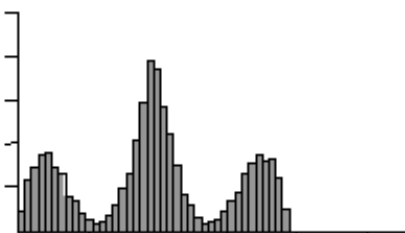
X



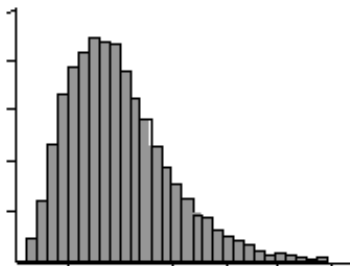
X



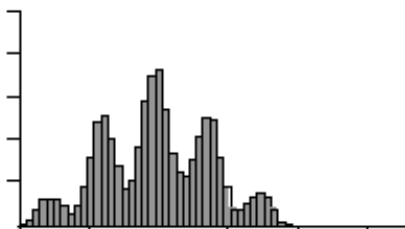
X
 $n=2$



X
 $n=2$



X
 $n=4$



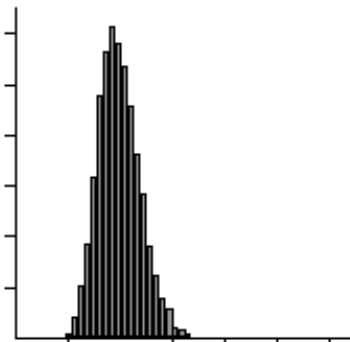
X
 $n=4$



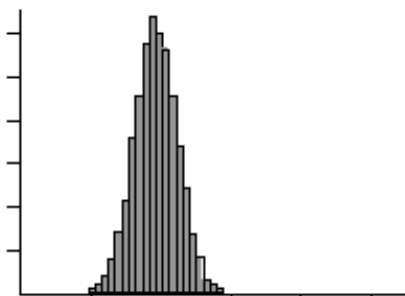
X
 $n=10$



X
 $n=10$



X
 $n=25$



X
 $n=25$

\bar{X} 的抽样分布

\bar{X} 的抽样分布

中心极限定理

(二) 从非正态(nonnormal)分布总体(均数为 μ , 方差为 σ)中随机抽样(每个样本的含量为 n), 可得无限多个样本, 每个样本计算样本均数, 则只要样本含量足够大($n > 50$), 样本均数也近似服从正态分布。

■ 样本均数的均数为 μ ;

■ 样本均数的标准差为 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ 。

3.标准误

standard error

- 抽样误差
- 中心极限定理
- 标准误
- 分布

标准误(standard error)

- 样本统计量的标准差称为标准误。样本均数的标准差称为均数的标准误。
- 均数的标准误表示样本均数的变异度。

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

- 当总体标准差未知时，用样本标准差代替，

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- 前者称为理论标准误，后者称为样本标准误。

与标准差的关系

区别

1、意义上

- 标准差描述个体值之间的变异，即观察值间的离散程度；
- 而标准误是描述统计量的抽样误差，即样本统计量和总体参数的接近程度；

2、用途上

- 标准差常用于表现观察值的波动范围；
- 标准误常表示抽样误差的大小，估计总体参数可信区间。

3、与样本含量

- 标准差是随着样本含量的增多，逐渐趋于稳定。
- 标准误是随着样本含量的增多，逐渐减少。

与标准差的关系

联系

- 首先，标准差和标准误都是变异指标，说明个体之间的变异用标准差，说明统计量之间的变异用标准误。
- 其次，当样本含量不变时，标准差大，标准误亦越大，均数的标准误与标准差成正比。

4. t 分布

t -distribution

- 抽样误差
- 中心极限定理
- 标准误
- 分布

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/298137065040006051>