

第 6 章 抽样理论

第 131 页 倒数第 3 行

统计量"的概率分布。第 7 章开始，就根据这些抽样统计量的概率分布，

第 132 页 第 6 行

(消除激烈的变动，见第 14 章)。

6.1 随机抽样

例题 6.1：假如总体有 5 个人，代号 A，B，C，D，E。写出下列所有的随机抽样结果。

1. 抽样数目 n = 2，有限总体随机抽样(抽样后不投返)。
2. 抽样数目 n = 2，随机抽样(抽样后投返)。
3. 抽样数目 n = 4，有限总体随机抽样(抽样后不投返)。

解答：1. 抽样数目 n = 2，有限总体随机抽样 (抽样后不投返)。

所有可能的随机抽样结果 ，有 5 ×4 = 20 种 ，以｛$X_1 = x_1$, $X_2 = x_2$｝表示：

｛A,B｝,｛A,C｝,｛A,D｝,｛A,E｝,｛B,A｝,｛B,C｝,｛B,D｝,｛B,E｝,｛C,A｝,｛C,B｝,｛C,D｝,｛C,E｝,｛D,A｝,｛D,B｝,｛D,C｝,｛D,E｝,｛E,A｝,｛E,B｝,｛E,C｝,｛E,D｝

如果不考虑其出现的排列次序，以组合方式，｛A,B｝=｛B,A｝，每种组合出现 2 次，所以可能的随机抽样结果，减为 $C_2^5 = 10$ 种：

｛A,B｝,｛A,C｝,｛A,D｝,｛A,E｝,｛B,C｝,｛B,D｝,｛B,E｝,｛C,D｝,｛C,E｝,｛E,D｝。

2. 抽样数目 n = 2，随机抽样(抽样后投返)。

所有可能的随机抽样有 $5^2 = 25$ 种：

｛A,A｝,｛A,B｝,｛A,C｝,｛A,D｝,｛A,E｝,｛B,A｝,｛B,B｝,｛B,C｝,｛B,D｝,｛B,E｝,｛C,A｝,｛C,B｝,｛C,C｝,｛C,D｝,｛C,E｝,｛D,A｝,｛D,B｝,｛D,C｝,｛D,D｝,｛D,E｝,｛E,A｝,｛E,B｝,｛E,C｝,｛E,D｝,｛E,E｝。

3. 抽样数目 n = 4，有限总体随机抽样(抽样后不投返)。

所有可能的随机抽样有 5×4×3×2 = 120 种：

如果不考虑其出现的排列次序，每种组合出现 4! = 24 次，所以可能的随机抽样减为 5 种：｛A,B,C,D｝,｛A,B,C,E｝,｛A,B,D,E｝,｛A,C,D,E｝,｛B,C,D,E｝。

6.4 中心极限定理 (Central Limit Theorem)

例题 6.3：总体概率分布函数 X_i 是贝努里分布 ，

$$P(\bar{x}) = \begin{cases} 0.4 & \text{if } x = 0 \\ 0.6 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

计算 抽样数目 n = 2,3,4 等的抽样平均 $\bar{X}$，并检查中心极限定理。

解答：因为 $X_i = 0$ 或 1，所以：如果 $\bar{x} = 0$，则 $x_1 = 0, x_2 = 0, \ldots, x_n = 0$。

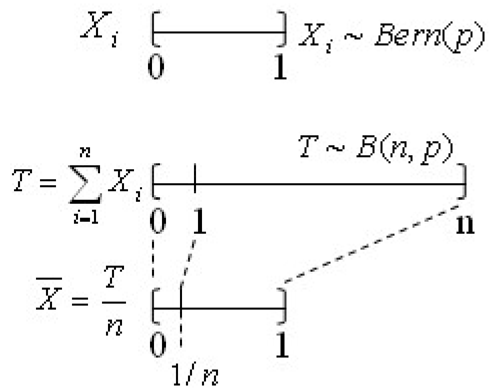
如果 $\bar{x} = 1$，则 $x_1 = 1, x_2 = 1, \ldots, x_n = 1$。

当抽样数目 n = 2，抽样平均 $\bar{X}$ 的分布如下：

$$P(\bar{x}) = \begin{cases} (0.4)^2 = 0.16 & \text{if } \bar{x} = 0 \\ 2(0.4)(0.6) = 0.48 & \text{if } \bar{x} = 1/2 \\ (0.6)^2 = 0.36 & \text{if } \bar{x} = 1 \end{cases}$$

当抽样数目 n = 3，抽样平均 $\bar{X}$ 的分布如下：

$$P(\bar{x}) = \begin{cases} (0.4)^3 = 0.064 & \text{if } \bar{x} = 0 \\ 3(0.4)^2(0.6) = 0.288 & \text{if } \bar{x} = 1/3 \\ 3(0.4)(0.6)^2 = 0.432 & \text{if } \bar{x} = 2/3 \\ (0.6)^3 = 0.216 & \text{if } \bar{x} = 1 \end{cases}$$



当抽样数目 n = 4，抽样平均 $\bar{X}$ 的分布如下：

$$P(\bar{x}) = \begin{cases} (0.4)^4 = 0.0256 & \text{if } \bar{x} = 0 \\ 4(0.4)^3(0.6) = 0.1536 & \text{if } \bar{x} = 1/4 \\ 6(0.4)^2(0.6)^2 = 0.3456 & \text{if } \bar{x} = 2/4 \\ 4(0.4)(0.6)^3 = 0.3456 & \text{if } \bar{x} = 3/4 \\ (0.6)^4 = 0.1296 & \text{if } \bar{x} = 1 \end{cases}$$

从以上 $\bar{X}$ 的概率分布，我们可以归纳出：如果总体分布是贝努里分布 $Bern(p)$，则对任何抽样数目 n，$\bar{X}$ 是一个二项分布 $B(n, p)$，只是将定义范围，从 0 到 n，缩到[0,1]。所以当 n 相当大(n>30)，$\bar{X}$ 变成连续型概率分布。

实际上 $n\bar{X}$ 是離散型二项分布，所以利用正态分布来作近似，求概率值时，要注意连续性修正，即计算 $P(\bar{X} \ge a)$ 改为 $P(\bar{X} \ge a - 0.5)$ 我们知道，如果 $np > 5$ 且 $n(1-p) > 5$，则二项分布趋近正态分布。

$$n\bar{X} \sim B(n, p) \cong N(np, np(1-p)), \quad E(n\bar{X}) = np, \quad E(\bar{X}) = p$$

$$V(n\bar{X}) = n^2 V(\bar{X}) = np(1-p), \quad V(\bar{X}) = \frac{p(1-p)}{n}，\text{所以：} \quad \bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

总体是贝努里分布，比例或成功率为 p，随机抽样(VSRS)样本量为 n，$X_i = 1$ 或 0，代表第 i 个样本是成功或失败，则成功的总数为 $T = \sum X_i$，抽样平均 $\bar{X} = \bar{P} = T/n$ 称为样本比例(sample proportion)。

根据中心极限定理当 $n \to \infty$，$\bar{P} \approx N\left(p, \frac{p(1-p)}{n}\right)$。

6.5 分层抽样

例题 6.4：一家公司想调查员工午餐的花费状况，计划从 1000 位员工抽出 40 人做样本。如果公司有 20%管理及行政人员，80%非管理的劳工。如何进行分层抽样？

解答：如果假设抽样回答率为 80%，为了得到 40 个有效样本要抽样 50 个。现在已经分成 2 层：管理人员及劳工，按照比率，管理人员抽样 10 人，劳工抽样 40 人。管理人员编号 1 到 200，随机抽样 10 个；劳工编号 1 到 800，随机抽样 40 个；合并以后就成为样本。

习题

1. 假设一总体为 4 个数字{2,4,6,8}，今以投返方式随机抽取两个数字 n=2 为样本：
   (1) 之抽样分布为何？
   (2) 写出总体的分布，并计算其平均数 μ 与标准差 σ。
   (3) 计算 $\overline{X}$ 抽样平均之期望值与标准差，证明分别为 μ 与$\frac{\sigma}{\sqrt{2}}$。
   (4) 如果是「不投返抽样」，则 $\overline{X}$ 之分布、期望值与标准差为何？
2. 关于上一题，假如我们所关心的不是统计量 $\overline{X}$ ，而是样本值全距 R={最大观察值-最小观察值}。例如，如果样本为{2,6}，则样本值全距为 6-2=4。
   (1) 求 R 之抽样分布为何 ？
   (2) 如果是「不投返抽样」，则 R 之抽样分布为何？
3. 假设一总体为正态分布，其平均数为 80，标准差为 10，有一随机抽样之样本量 n=9：
   (1) $\overline{X}$ 之分布为何？这是正确的或近似的？求 $\overline{X}$ 的平均数与标准差。
   (2) 求 $\overline{X}$ 在 76 至 84 间之概率为何？
4. 一个骰子掷 60 次，计算其 60 次的总和 $T=\sum_{i=1}^{60} X_i$
   (1) T 期望值与方差是多少 ？
   (2) $P(T \le 200)$是多少 ？
5. 假设某河床散布之石块体积为正态分布，平均数为 121，标准差为 3.2。今随机选 16 颗石头为样本，令 $\overline{X}$ 表示样本之平均体积。
   (1) 求 $\overline{X}$ 之分布为何？
   (2) $\overline{X}$ 小于 120 之概率为何 ？
   (3) 该河床中体积小于 120 之石子占全体之比例为何 ？
6. 假设学生完成上课登记所花时间平均数为 94 分钟，标准差为 10 分钟，今有 81 位学生随机之样本：
   (1) 求 $\overline{X}$ 的平均数与标准差。
   (2) $\overline{X}$ 之分布有何特性？
   (3) $P[\overline{X} > 96]$
   (4) $P[92.3 < \overline{X} < 96]$
   (5) $P[\overline{X} < 95]$
7. 假设某堆行李平均重量为 55 磅，标准差为 7 磅。如果随机抽取 40 件行李为样本，样本平均重量 $\overline{X}$ 位于 54 至 56 磅间之概率为何？
8. 假设某大学某系学生每月看电影次数呈下列之分布：

数目	概率值 P(x)
0	0.5
1	0.4
2	0.1

   (1) 随机抽样 n=2，试求 $\overline{X}$ 之概率分布、平均数与标准差。
   (2) 如果总体只有 20 人，当 n = 2 时，试求「有限总体抽样平均」$\overline{X}$ 之概率分布、平均数与标准差。

9. 某果汁供货商提供瓶装果汁给超级市场贩卖、罐装果汁给便利商店贩卖。此供货商想从这两种管道的销售资料中统计分析，了解究竟哪种包装的果汁较受消费者欢迎，请根据这个统计分析回答以下的问题。
   (1) 请问这个统计分析是观察性还是实验性？请解释理由。
   (2) 由于这个统计分析只有此供货商本身的出货资料，而且两种销售管道贩卖不同包装的果汁，无法第一手了解消费者的喜好，请建议更好的方法来分析。
10. 从平均年龄为 75 岁与标准差为 6 岁的正态老年人的总体中随机抽样 50 位老年人，请根据这个随机样本回答以下的问题。
    (1) 请问这个随机样本平均年龄的样本型态？请解释。
    (2) 请计算这个随机样本平均年龄的样本平均数与标准误。
    (3) 请问老年人样本平均年龄超过 73 岁的概率？
    (4) 请问老年人样本平均年龄少于 75 岁的概率？
    (5) 请问随机连续抽样两位老年人，平均年龄都超过 73 岁的概率？
11. 根据统计台湾成年人每星期运动的时间，是一平均为 4 小时与标准差为 1.5 小时的正态分布，若从台湾成年人的总体中随机抽样 100 位，请根据这个随机样本回答以下的问题。
    (1) 请问这个随机样本台湾成年人平均每星期运动时间的样本型态？请解释。
    (2) 请计算这个随机样本台湾成年人平均每星期运动时间的样本平均数与标准误。
    (3) 请问台湾成年人平均每星期运动时间超过 5 小时的概率？
    (4) 请问台湾成年人平均每星期运动时间少于 3 小时的概率？
    (5) 请问随机连续抽样 4 位台湾成年人，平均每星期运动时间都超过 5 小时的概率？
12. 根据统计台湾某地区最畅销的报纸每天卖掉份数，是一平均为 2000 份与标准差为 500 份的正态分布，若从此地区报纸销售的总体中连续抽样 60 天，请根据这个样本回答以下的问题。
    (1) 请问此地区报纸每天销售平均份数的样本型态？请解释。
    (2) 请计算此地区报纸每天销售平均份数的样本平均数与标准误。
    (3) 请问此地区报纸每天销售平均份数介于 1900 份与 2300 份的概率？
    (4) 请问此地区报纸每天销售平均份数超过 2500 份的概率？
    (5) 请问随机连续抽样 4 天，此地区报纸每天销售份数都超过 2300 份的概率？
13. 根据某民意调查公司的实时民调，若现在选择台湾的代表花种，约有 65%的人会选择百合花，为了解是否百合花为大家选择，若从台湾成年人的总体中随机抽样 1000 位，请根据这个样本回答以下的问题。
    (1) 假设此民意调查公司的实时民调是正确的，请问此台湾成年人选择百合花为台湾代表花种的样本分布型态？请解释。
    (2) 请计算台湾成年人选择百合花为台湾代表花种的样本平均数与标准误。
    (3) 请问台湾成年人选择百合花为台湾代表花种的人数介于 350 人与 680 人的概率？
    (4) 请问台湾成年人选择百合花为台湾代表花种的人数超过 500 人的概率？
    (5) 请问随机连续抽样 20 人，选择百合花为台湾代表花种的人数超过 15 人的概率？
14. 某网络咖啡厅想了解其顾客有多少是 21 岁以下地年轻人，因此抽样了共 500 位顾客发现其中有 375 位是 21 岁以下地年轻人，但是根据政府普查发现此地区应有约 68%的人口是 21 岁以下地年轻人，请根据这个样本回答以下的问题。

(1) 请问此网络咖啡厅 21 岁以下地年轻人顾客比例的样本分布型态？请解释。

(2) 请计算此网络咖啡厅 21 岁以下地年轻人顾客比例的样本平均数与标准误。

(3) 请问此网络咖啡厅 21 岁以下地年轻人顾客比例介于 3%此地区 21 岁以下地年轻人口的概率？

(4) 请问此网络咖啡厅 21 岁以下地年轻人顾客比例介于 5%此地区 21 岁以下地年轻人口的概率？

(5) 请问随机连续抽样 20 人，此网络咖啡厅 21 岁以下地年轻人顾客的人数超过 10 人的概率？

15. 假设研究发现服务业女性员工起薪为一平均月薪 NT$23000 且标准差为 NT$8200 的右偏分布，而男性员工起薪为一平均月薪 NT$25000 且标准差为 NT$9600 的右偏分布，为研究此差异抽样了共 50 位女性员工与 45 位男性员工，请根据这个样本回答以下的问题。

(1) 请问服务业女性员工起薪与男性员工起薪差异的样本分布型态？请解释。

(2) 请计算服务业女性员工起薪与男性员工起薪差异的样本平均数与标准误。

(3) 请问服务业女性员工起薪高于男性员工起薪的概率？

(4) 请问服务业女性员工起薪小于等于男性员工起薪的概率？

(5) 请问服务业女性员工起薪高于男性员工起薪至少 NT$5000 的概率？