

第 8 章

抽樣分配

定義 8-0-1：估計量 (estimator)

作為估計用的樣本統計量稱為估計量。

定義 8-0-2：檢定量 (tester)

作為假設檢定用的樣本統計量稱為檢定量。

定義 8-1-1：抽樣分配 (sampling distribution)

樣本統計量的機率分配稱為抽樣分配。

定理 8-2-1：母群體隨機變數 X 的期望值為 μ (或 μ_x)，變異數為 σ^2 (或 σ_x^2)，換言之即 $X \sim (\mu_x, \sigma_x^2)$ 。 $x_1, x_2, \dots, x_n$ 為 n 個來自這個母群體的隨機樣本，

$\bar{X} = \frac{x_1 + \dots + x_n}{n}$ 為這 n 個樣本的樣本平均數，則

$$(1) E(\bar{X}) = \mu_x$$

$$(2) V(\bar{X}) = \frac{\sigma_x^2}{n} = \frac{V(X)}{n}$$

定理 8-2-2：有限母群體，不放回抽樣方式下定理 8-2-1 的修正

有限母群體共有 N 個元素，其隨機變數之期望值為 μ (或 μ_x)，變異數為 σ^2 (或 σ_x^2)， $x_1, \dots, x_n$ 為採不放回方式自母群體抽樣得到的 n 個樣

本值 ($n < N$)， $\bar{X}$ 為這 n 個樣本的樣本平均數，則

$$(1) E(\bar{X}) = \mu_x$$

$$(2) V(\bar{X}) = \frac{\sigma_x^2}{n} \cdot \left(\frac{N-n}{N-1} \right)$$

定理 8-3-1：中央極限定理 (the central limit theorem)

母群體隨機變數 X 的期望值為 μ ，變異數為 σ^2 ，換言之， $X \sim (\mu, \sigma^2)$ ， $x_1, \dots, x_n$ 為抽樣自此母群體的 n 個隨機樣本，當抽樣數目 n 足夠大時

(1) $\bar{X}$ 為近似常態分配，且其平均數為 μ ，變異數為 $\frac{\sigma^2}{n}$ 。換言之，

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

(2) $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ 為近似標準常態分配 $N(0, 1)$

定理 8-3-2：母群體隨機變數 X 的期望值為 μ ，變異數為 σ^2 。換言之， $X \sim (\mu, \sigma^2)$ 。

$x_1, \dots, x_n$ 為抽樣自此母群體的 n 個隨機樣本，當抽樣數目 n 足夠大時，

樣本總和 $Y = \sum_{i=1}^n X_i$ 的抽樣分配為一個平均數為 $n\mu$ ，變異數為 $n\sigma^2$ 的近似常態分配，也就是

$$(1) Y = \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

$$(2) \frac{Y - n\mu}{\sqrt{n}\sigma} \approx N(0, 1)$$

定理 8-3-3：常態母群體的樣本總和及樣本平均數之抽樣分配

母群體隨機變數 X 的期望值 μ ，變異數 σ^2 的常態分配，也就是 $X \sim N(\mu, \sigma^2)$ 。 $x_1, \dots, x_n$ 為抽樣自此母群體的 n 個隨機樣本，則

$$(1) Y = \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$

$$(2) \frac{Y - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$

$$(3) \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$(4) \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

注意：上述(4)項結論不受 n 大小的限制。

定理 8A-1-1：大數法則 (law of large number)

$X \sim (\mu, \sigma^2)$ 則

$$P(|\bar{X} - \mu| < c) \geq 1 - \frac{\sigma^2}{nc^2}$$

定理 8A-2-1：常態分配的再生性 (regeneration)

$X_i, i=1, \dots, n$ 為抽樣自 $N(\mu_i, \sigma_i^2)$ 的隨機樣本，則

$\sum_{i=1}^n a_i X_i = a_1 X_1 + \dots + a_n X_n$ 的抽樣分配為

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

定理 8A-3-1： $Z \sim N(0, 1)$ 則 Z^2 的機率分配為自由度 1 的 χ^2 分配，也就是 $Z^2 \sim \chi^2(1)$

定理 8A-3-2： χ^2 分配的再生性 (regeneration)

$X_1, \dots, X_n$ 為抽樣自 $\chi^2(v_i)$ 的隨機樣本，則

$\sum_{i=1}^n X_i = X_1 + \dots + X_n$ 的抽樣分配為

$$\sum_{i=1}^n X_i \sim \chi^2\left(\sum_{i=1}^n v_i\right)$$

定理 8A-3-3： $X_1, \dots, X_i, \dots, X_n$ 為抽樣自 $N(\mu_i, \sigma_i^2)$ 的隨機樣本，

則 $\sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$ 的抽樣分配為

$$\sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2(n)$$

定理 8A-3-4： $X_i, i=1, \dots, n$ 為抽樣自 $N(\mu, \sigma^2)$ 的隨機樣本，則

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \text{ 的抽樣分配為}$$

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

定理 8A-3-5： $X \sim \chi^2(v)$ 且 $Z \sim N(0, 1)$ 則

$$\frac{Z}{\sqrt{\frac{X}{v}}} \sim t(v)$$

定理 8A-3-6： $X_1, \dots, X_n$ 為抽樣自 $N(\mu, \sigma^2)$ 的隨機樣本，當 σ^2 未知而以 S^2 替代，則

$$\frac{\frac{\bar{X} - \mu}{S}}{\sqrt{n}} \text{ 的抽樣分配為}$$

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

8.1 從五個數字 1, 3, 4, 8, 9 所組成的母群體隨機抽取 3 個樣本，以 $\bar{X}$ 表示樣本平均數， m 表示樣本中位數

- (1) 計算母群體的期望值與變異數。
- (2) 推導 $\bar{X}$ 的抽樣分配，並計算其期望值與變異數。
- (3) 推導 m 的抽樣分配，並計算其期望值與變異數。
- (4) 比較 $\bar{X}$ 與 m 這兩個估計量用於估計母群體平均數 μ 時之效果。

解：

以 X 表示母群體隨機變數， μ ， σ^2 分別表示它的期望值與變異數，則

$$(1) \mu = E(X) = (1 + 3 + 4 + 8 + 9) \times \frac{1}{5} = 5$$

$$\sigma^2 = E(X^2) - E^2(X) = (1^2 + 3^2 + 4^2 + 8^2 + 9^2) \times \frac{1}{5} - 5^2 = 9.2$$

(2) 抽樣結果共有 10 種 $\left(\binom{5}{3} = 10\right)$ 可能，分別計算 $\bar{X}$ 及 m 如下表

樣本	$\bar{X}$	m
(1,3,4)	2.67	3
(1,3,8)	4	3
(1,3,9)	4.33	3
(1,4,8)	4.33	4
(1,4,9)	4.67	4
(1,8,9)	6	8
(3,4,8)	5	4
(3,4,9)	5.33	4
(3,8,9)	6.67	8
(4,8,9)	7	8

每個樣本發生的機會均等 $\left(\frac{1}{10}\right)$ ，所以， $\bar{X}$ 的抽樣分配為

134 • 統計學習題解答

$\bar{X}$	2.67	4	4.33	4.67	5	5.33	6	6.67	7
$f(\bar{x})$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$

所以 $\bar{X}$ 的期望值與變異數為

$$\begin{aligned}
 E(\bar{X}) &= 4.33 \times \frac{2}{10} + (2.67 + 4 + 4.67 + 5 + 5.33 + 6 + 6.67 + 7) \times \frac{1}{10} \\
 &= 5 \\
 V(\bar{X}) &= E(\bar{X}^2) - E^2(\bar{X}) \\
 &= \left[4.33^2 \times \frac{2}{10} + (2.67^2 + 4^2 + 4.67^2 + 5^2 + 5.33^2 + 6^2 + 6.67^2 + 7^2) \times \frac{1}{10} \right] - 25 \\
 &= 1.5332
 \end{aligned}$$

(3) 同理， m 的抽樣分配為

m	3	4	8
$f(m)$	$\frac{3}{10}$	$\frac{4}{10}$	$\frac{3}{10}$

所以， m 的期望值與變異數為

$$\begin{aligned}
 E(m) &= 3 \times \frac{3}{10} + 4 \times \frac{4}{10} + 8 \times \frac{3}{10} = 4.9 \\
 V(m) &= E(m^2) - E^2(m) = \left(3^2 \times \frac{3}{10} + 4^2 \times \frac{4}{10} + 8^2 \times \frac{3}{10} \right) - (4.9)^2 = 4.29
 \end{aligned}$$

(4) $\bar{X}$ 為 μ 的不偏估計量 (unbiased estimator)，因為它的期望值與 μ 相等，而 m 則不是一個不偏估計量。此外， $\bar{X}$ 的變異數較 m 為小，這表示 $\bar{X}$ 比 m 穩定。

8.2 母群體隨機變數 X 呈現出右偏 (skewed to right)，且其平均數為 30，標準差為 26，從這個母群體中抽取 169 個樣本， $\bar{X}$ 為這 169 個樣本值的平均數，則

(1) $\bar{X}$ 的機率分配為何？

(2) 計算 $P(\bar{X} \geq 34) = ?$

解：

(1) 由於抽樣數目 169，根據中央極限定理，樣本平均數 $\bar{X}$ 的分配近似常態分配，且

$$\bar{X} \sim N\left(30, \frac{(26)^2}{169}\right)$$

$$\begin{aligned} (2) P(\bar{X} \geq 34) &= P\left(\frac{\bar{X} - 30}{\frac{26}{\sqrt{169}}} \geq \frac{34 - 30}{\frac{26}{\sqrt{169}}}\right) \\ &= P(Z \geq 2) \\ &= 0.0228 \end{aligned}$$

8.3 無限母群體隨機變數 X 的期望值 $\mu = 30$ ，變異數 $\sigma^2 = 10$ ，從這個母群體中隨機取樣 $x_1, \dots, x_{20}$ (抽樣數 $n = 20$)，

(1) 樣本平均數 $\bar{X}$ 的期望值 $E(\bar{X}) = ?$

(2) 樣本平均數 $\bar{X}$ 的變異數 $V(\bar{X}) = ?$

解：

由定理 8-2-1，得知

$$(1) E(\bar{X}) = \mu = 30$$

$$(2) V(\bar{X}) = \frac{\sigma^2}{n} = \frac{10}{20} = \frac{1}{2}$$

8.4 常態母群體之期望值為 10，變異數為 441。從這個母群體中隨機抽樣 $x_1, \dots, x_{49}$ ， $\bar{X}$ 為樣本平均，則

(1) 計算 $P(\bar{X} \geq 13) = ?$

(2) 計算 $P(4 \leq \bar{X} \leq 13) = ?$

解：

由定理 8-3-3，得知，

$$\bar{X} \sim N\left(10, \frac{441}{49}\right)$$

$$\begin{aligned}(1) P(\bar{X} \geq 13) &= P\left(\frac{\bar{X} - 10}{\sqrt{\frac{441}{49}}} \geq \frac{13 - 10}{\sqrt{\frac{441}{49}}}\right) \\ &= P(Z \geq 1) \\ &= 0.1587\end{aligned}$$

$$\begin{aligned}(2) P(4 \leq \bar{X} \leq 13) &= P\left(\frac{4 - 10}{\sqrt{\frac{441}{49}}} \leq \frac{\bar{X} - 10}{\sqrt{\frac{441}{49}}} \leq \frac{13 - 10}{\sqrt{\frac{441}{49}}}\right) \\ &= P(-2 \leq Z \leq 1) \\ &= 0.8185\end{aligned}$$

8.5 某化工廠所製造的 5 公斤桶裝油漆中，某種有毒性揮發物質的平均含量為 4.0 公克，標準差為 1.3 公克（並假設它為常態分配）

- (1) 這家化工廠所製造的桶裝油漆中，這種有毒性揮發物含量超過 5 公克的比例有多少？
- (2) 隨機抽樣 100 桶油漆，其樣本平均含量超過 4.5 公克的機率有多少？
- (3) 如果檢驗單位在某次檢驗中，測得 100 桶的樣本平均含量為 4.5 公克，你認為該化工廠的生產線可能發生了什麼問題？

解：

令 X 表示該化工廠所製造 5 公斤桶裝油漆中有毒揮發物的含量，則由題意得知

$$X \sim N(4.0, (1.3)^2)$$

$$\begin{aligned}(1) P(\text{有毒揮發物超過 } 5 \text{ 公克}) &= P(X > 5) \\ &= P\left(\frac{x - 4.0}{1.3} > \frac{5 - 4.0}{1.3}\right) \\ &= P(Z > 0.77) \\ &= 0.2207\end{aligned}$$

所以，這家化工廠所生產的油漆中，有 22.07%有毒揮發物超過 5 公克。

(2)該化工廠所生產油漆中有毒揮發物的含量為常態分配，根據定理 8-3-3， $\bar{X}$ 為常態分配，也就是

$$\begin{aligned}\bar{X} &\sim N\left(4.0, \frac{(1.3)^2}{100}\right) \\ P(100 \text{ 桶樣本平均值超過 } 4.5) \\ &= P(\bar{X} \geq 4.5) \\ &= P\left(\frac{\bar{X} - 4.0}{\sqrt{\frac{(1.3)^2}{100}}} \geq \frac{4.5 - 4.0}{\sqrt{\frac{(1.3)^2}{100}}}\right) \\ &= P(Z \geq 3.85) \\ &= 0.0001\end{aligned}$$

(3)從(2)的計算中得知，100 桶樣本平均含量超過 4.5 是極為罕見的結果（機率為 0.0001），而檢驗單位竟然檢驗出這個極不可能看到的數值，則我們認為該化工廠的生產線極可能出了某種問題，使得它原來的生產規格（平均含量 4.0，標準差 1.3）改變了。換句話說。生產線所生產的有毒物平均含量增加了。

8.6 依衛生署的相關規定，罐頭食品某種可能致癌物質的含量不可超過 2.0 公克，某家食品製造商所生產的罐頭食品中，這種物質的含量呈現常態分配，其平均數為 1.25 公克，標準差為 0.5 公克

(1)這個食品製造商所生產的罐頭中有多少百分比不符合衛生署的規定。

(2)假設衛生署定期抽檢的程序如下：抽樣 25 個罐頭，並計算 25 個可能致癌物含量的平均數，如果樣本平均值超過 1.6 公克則廠商將受到嚴厲的處罰。則這家廠商遭到处罰的機率是多少。此外，你認為這個檢驗規則是否合理。

解：

令 X 表示罐頭中可能致癌物的含量，則

$$X \sim N(1.25, (0.5)^2)$$

$$\begin{aligned}
 (1) P(\text{不合衛生署規定}) &= P(X \geq 2) \\
 &= P\left(\frac{X - 1.25}{0.5} \geq \frac{2 - 1.25}{0.5}\right) \\
 &= P(Z \geq 1.5) \\
 &= 0.0668
 \end{aligned}$$

(2) 令 $\bar{X}$ 表示 25 個樣本中可能致癌物質的平均含量，則

$$\begin{aligned}
 \bar{X} &\sim N\left(1.25, \frac{(0.5)^2}{25}\right) \\
 P(\text{食品製造商遭衛生署嚴厲處罰}) \\
 &= P(\bar{X} \geq 1.6) \\
 &= P\left(\frac{\bar{X} - 1.25}{\frac{0.5}{5}} \geq \frac{1.6 - 1.25}{\frac{0.5}{5}}\right) \\
 &= P(Z \geq 3.5) \\
 &= 0.0002
 \end{aligned}$$

從(1)、(2)的計算結果顯示，這家廠商生產出大約 7% 的不合規定的產品，但卻只承擔了 0.02% 的處罰風險，所以，檢驗規則應該再嚴一些才合理。

8.7 假設顧客在銀行等候線等待服務的時間為常態分配，其平均數及標準差分別為 100 秒及 10 秒。顧客可接受的等待時間極大值也是常態分配，其平均數及標準差分別為 95 秒及 8 秒。當顧客等待的時間超過他（她）可接受的極限值時，他（她）會有不滿意的感覺，則

(1) 當你到這家銀行，你的等待時間不超過 95 秒的機率是多少？

(2) 整體來說，有多少比例的顧客會有不滿意的感覺。

解：

(1) 令 X 表示等待時間，則

$$\begin{aligned}
 X &\sim N(100, 10^2) \\
 P(\text{等待時間不超過 95 秒}) &= P(X \leq 95)
 \end{aligned}$$

$$\begin{aligned}
 &= P\left(\frac{X-100}{10} \leq \frac{95-100}{10}\right) \\
 &= P(Z \leq -0.5) \\
 &= 0.3085
 \end{aligned}$$

這表示，你到達銀行時，有 30.85% 的機會，等待時間不超過 95 秒，但注意，這並非表示將只有 30.85% 的顧客會有滿意的感覺，因為滿意與否是個人的可接受等待時間極限值而異的。

(2) 令 Y 表示顧客可接受的等待時間極限值，則

$$Y \sim N(95, 8^2)$$

我們可以很合理的假設 X 與 Y 互相獨立，所以

$$X - Y \sim N(100 - 95, 10^2 + 8^2)$$

當 $X > Y$ (或 $X - Y > 0$) 時，顧客將有不滿意的感覺，所以

$$\begin{aligned}
 P(\text{整體來說，顧客不滿意}) &= P(X - Y > 0) \\
 &= P\left(\frac{(X - Y) - 5}{\sqrt{164}} \geq \frac{0 - 5}{\sqrt{164}}\right) \\
 &= P(Z \geq -0.39) \\
 &= 0.6517
 \end{aligned}$$

這表示，整體來說將有 65.17% 的顧客不滿意。

8.8 某餐飲店統計資料顯示，顧客消費金額的平均值為 100 元，標準差為 18 元，則

- (1) 36 位顧客的平均消費金額超過 97 元的機率是多少？
- (2) 100 位顧客的營業總金額超過 10,500 元的機率是多少？
- (3) 該餐飲店若要使每日的營業總金額超過 12000 元的機率大於 95%，則每日至少要有多少位顧客光臨。

解：

- (1) 令 X 表示每位顧客的消費金額，則

$$X \sim (100, 18^2)$$

根據中央極限定理，36 位顧客的平均消費金額 $\bar{X}$ 為近似常態分配。

$$\begin{aligned}\bar{X} &\sim N\left(100, \frac{18^2}{36}\right) \\ P(36 \text{ 位顧客平均消費額超過 } 97 \text{ 元}) \\ &= P(\bar{X} > 97) \\ &= P\left(\frac{\bar{X} - 100}{3} > \frac{97 - 100}{3}\right) \\ &= P(Z > -1) \\ &= 0.8413\end{aligned}$$

(2) 令 $X_1, X_2, \dots, X_{100}$ 表示 100 位顧客的消費金額， $Y = \sum_{i=1}^{100} X_i$ 表示這 100 位顧客的總消費金額，根據定理 8-3-2， Y 的機率分配為近似常態分配，也就是

$$\begin{aligned}Y &\sim N(10,000, 100 \times 18^2) \\ P(100 \text{ 位顧客總消費金額超過 } 10500 \text{ 元}) \\ &= P(Y > 10,500) \\ &= P\left(\frac{Y - 10,000}{180} > \frac{10,500 - 10,000}{180}\right) \\ &= P(Z > 2.78) \\ &= 0.0027\end{aligned}$$

(3) 令 Y_n 表示 n 位顧客的總消費金額，也就是 $Y_n = \sum_{i=1}^n X_i$ ，根據定理 8-3-2

$$Y \sim N(100n, n \times 18^2)$$

所以，要使得每日營業金額超過 12000 元的機率大於 0.95，也就是

$$P(Y > 12000) > 0.95$$

$$P\left(\frac{Y-100n}{18\sqrt{n}} > \frac{12,000-100n}{18\sqrt{n}}\right) > 0.95$$

查表得知，滿足上式的 Z 值為 -1.645 ，所以

$$\begin{aligned}\frac{12,000-100n}{18\sqrt{n}} &< -1.645 \\ 29.61\sqrt{n} &< 100n-12000 \\ 876.75n &< 10000n^2-2400000n+144000000 \\ 10,000n^2-2400876.7n+144000000 &> 0 \\ n &> \frac{2400876.7+\sqrt{(2400876.7)^2-4\times 10000\times 144000000}}{2\times 10000} \\ n &> 123.2876\end{aligned}$$

所以，為了確保每日營業額超過 12000 元的機率大於 0.95，該餐飲店的每日顧客必需超過 124 人。

8.9 統計資料顯示，所有上市公司股票的年投資報酬率的平均值為 9%，標準差為 18%，某投資理財雜誌社隨機取樣 36 位訂戶，計算出這 36 位訂戶投資上市公司股票的平均年報酬率為 6%

(1) 在雜誌社訂戶年投資報酬率分配與所有上市公司年投資報酬率分配相同的假設下，雜誌社 36 位訂戶年投資報酬率低於 6% 的機率是多少？

(2) 根據上述資料，你（妳）是否認為該雜誌社訂戶的投資績效不佳。

解：

(1) 令 $X_1, \dots, X_{36}$ 表示 36 位訂戶的年投資報酬率，因為這些訂戶的年投資報酬率分配與上市公司股票年投資報酬率相同，表示，它們有相同的平均數與標準差，所以

$$X_i \sim (0.09, (0.18)^2), \quad i=1, 2, \dots, 36$$

令 $\bar{X}$ 表示這 36 位雜誌訂戶的年投資報酬率平均值，根據中央極限定理， $\bar{X}$ 為近似常態分配

$$\bar{X} \sim N\left(0.09, \frac{(0.18)^2}{36}\right)$$

所以，這 36 位雜誌訂戶的年投資報酬率平均值低於 6% 的機率為

$$\begin{aligned} & P(36 \text{ 位雜誌訂戶年平均投資報酬率低於 } 6\%) \\ &= P(\bar{X} < 0.06) \\ &= P\left(\frac{\bar{X} - 0.09}{0.03} < \frac{0.06 - 0.09}{0.03}\right) \\ &= P(Z < -1) \\ &= 0.1587 \end{aligned}$$

(2) 根據(1)的計算，只有 15.87% 機率使得 36 位訂戶的年投資報酬率平均值低於 6%，但是它卻發生了，如果沒有抽樣偏差（這 36 位的投資狀況正確反映了所有該雜誌的訂戶），這顯示該雜誌訂戶的投資策略或投資標的的選擇不是很好。

（註）讀者學過統計檢定後，你（妳）將有更精準更嚴謹的判斷。