工程統計-第四章

常態分布與應用

第四章 ： 常態分布與應用

- 4-1 隨機變數與機率分配
- 4-2 常態分配
- 4-3 隨機抽樣
- 4-4 常態分配
  - 水文頻率分析
  - 平均數控制

4-1 隨機變數與機率分配

將隨機試驗所有可能結果轉為數值(事件機率)所使用的變數稱為<u>隨機變數(Random Variable)</u>。

◎隨機變數的函數

1.間斷式機率函數

(Discrete Probability Distribution

2.連續式機率函數

(Continuous Random Variable)

◎連續型隨機變數

連續分配之機率函數稱為機率密度函數(Probability Density Function)，以 $f(x)$ 函數表示。

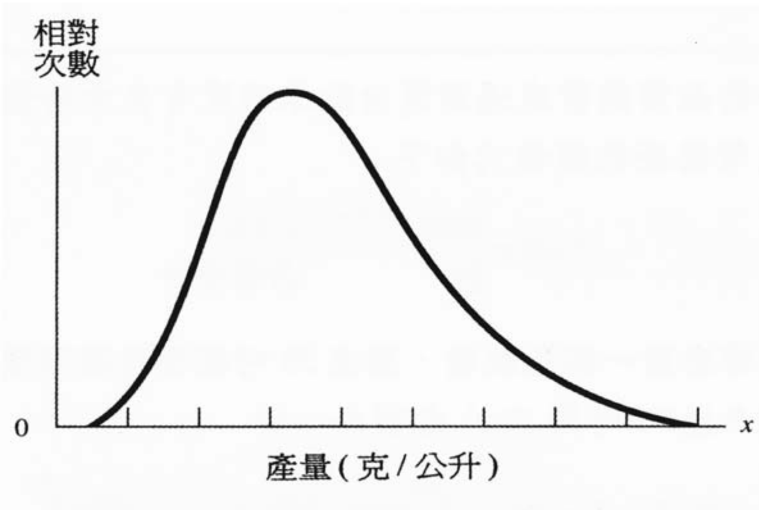


圖 7-3 化學有益成分產量的母體次數分配圖

◎機率密度函數

對於連續型隨機變數 X 之所有可能值範圍而言，

其機率密度函數 $f(x) \geq 0$ 。

而介於 a 與 b 間的面積為

$$\Pr[a \leq X \leq b] = \int_a^b f(x)dx$$

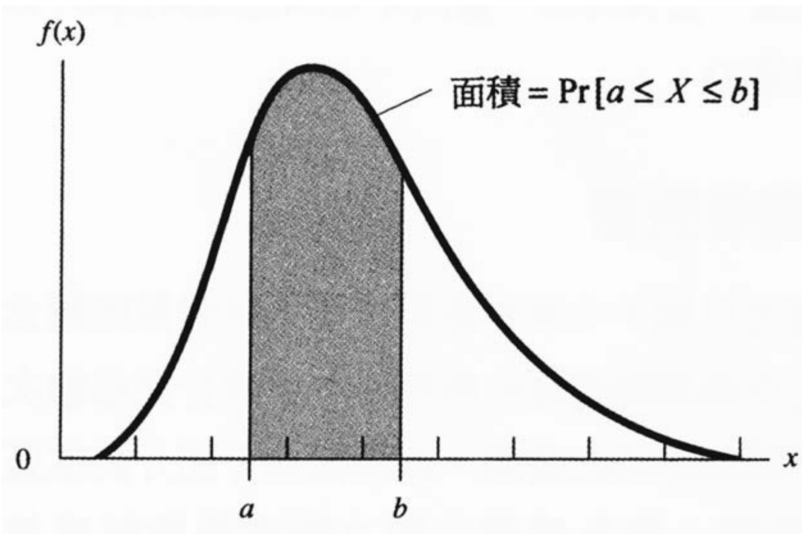


圖 7-4　有益成分產量之機率密度函數曲線

4-2 常態分配

常態分配(Normal Distribution)又可稱為**高斯分配**(Gaussian Distribution)。

◎常態分配與母體次數分配曲線

常態分配為統計學最重要的分配之一，當樣本數足夠大(一般而言，$n \geq 30$)，則間斷型機率分配或連續型機率分配，可以用常態分配來逼近。常態分配圖形為一**鐘型**(Bell Shape)對稱的曲線。

母體具平均數μ與標準差σ之隨機變數 X 的常態分配的機率密度函數為：

Probability Density Function

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

Probability Distribution Function

$$F(x) = \Pr[X \leq x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dy$$

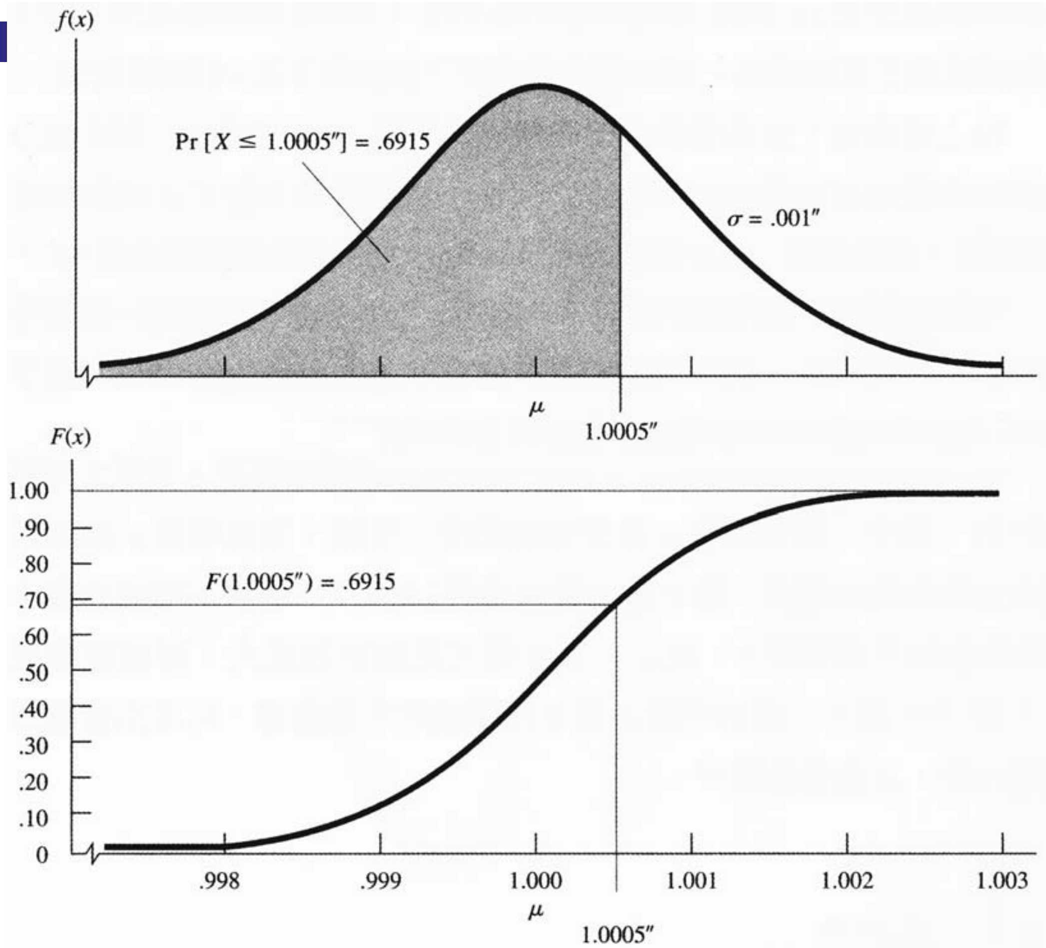


圖 7-10 常態機率密度函數（上圖）與常態機率分配函數（下圖）之曲線

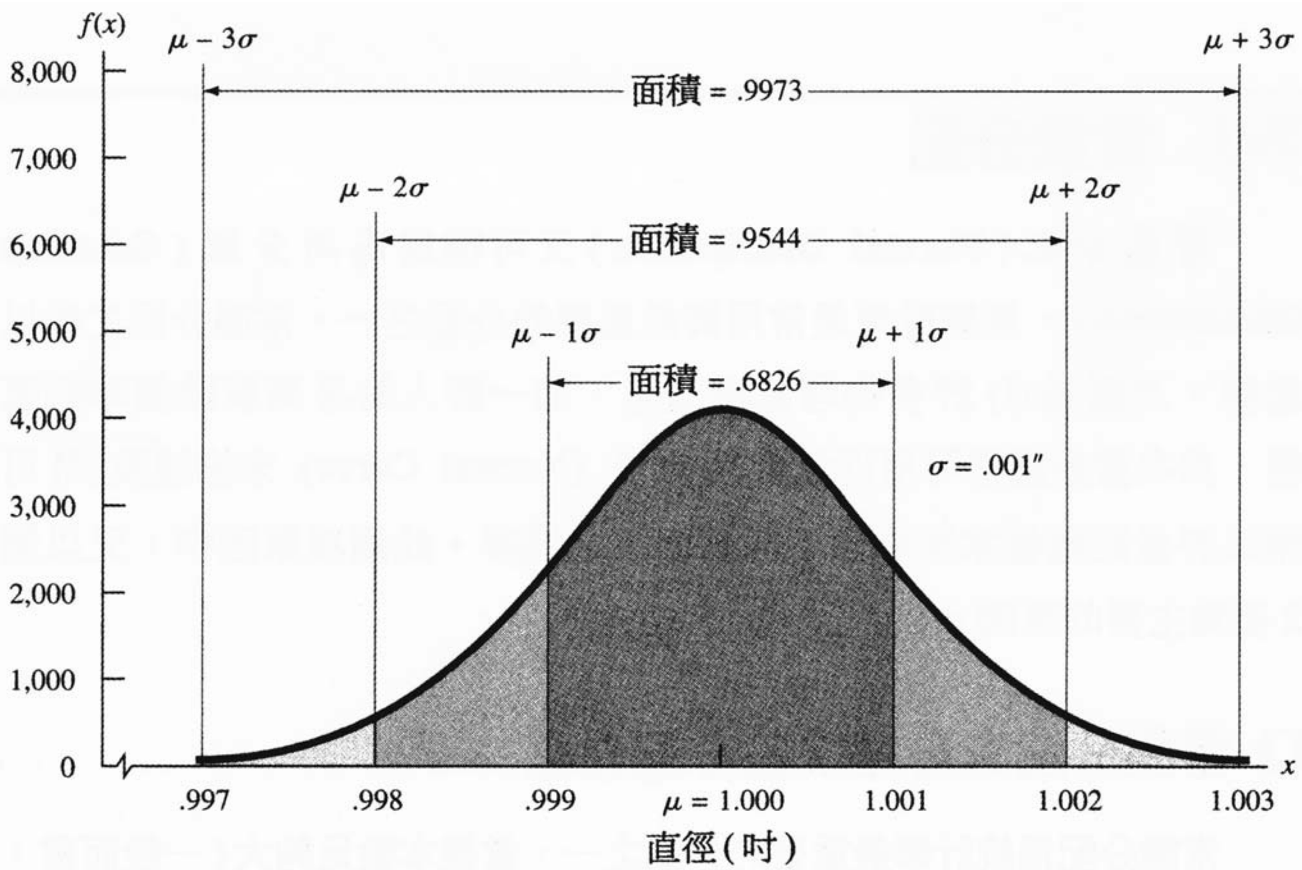
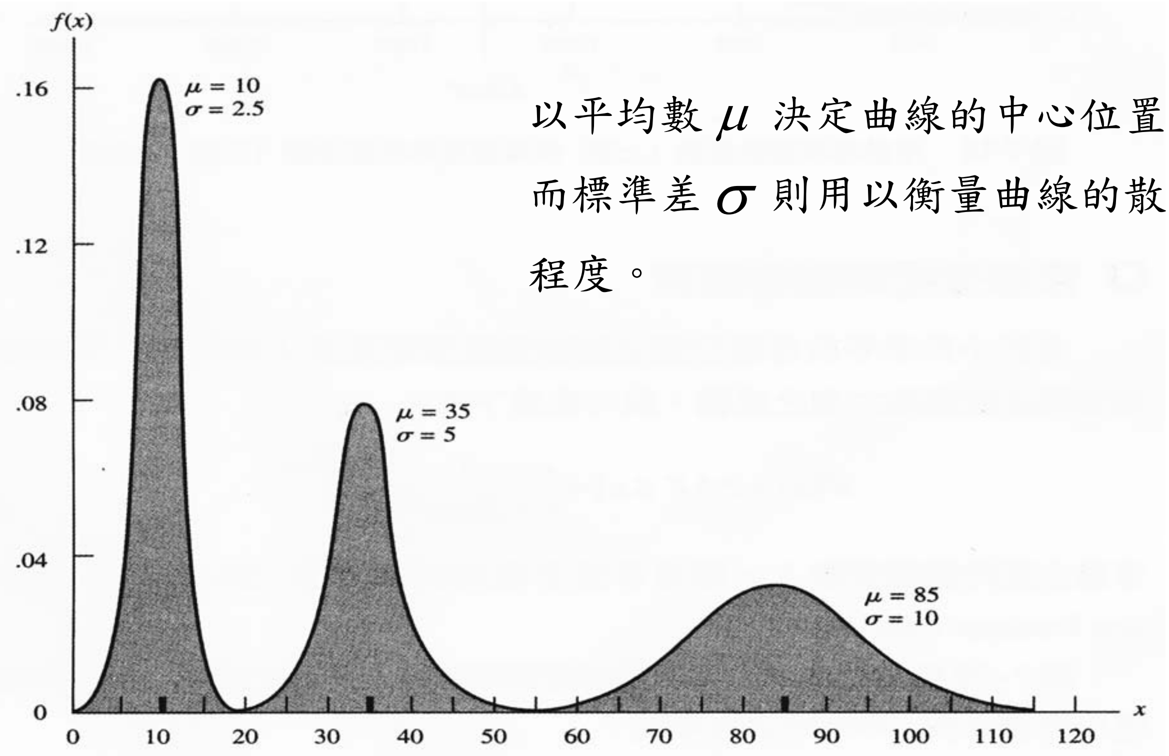


圖 7-8 軸承直徑 1 吋之常態曲線



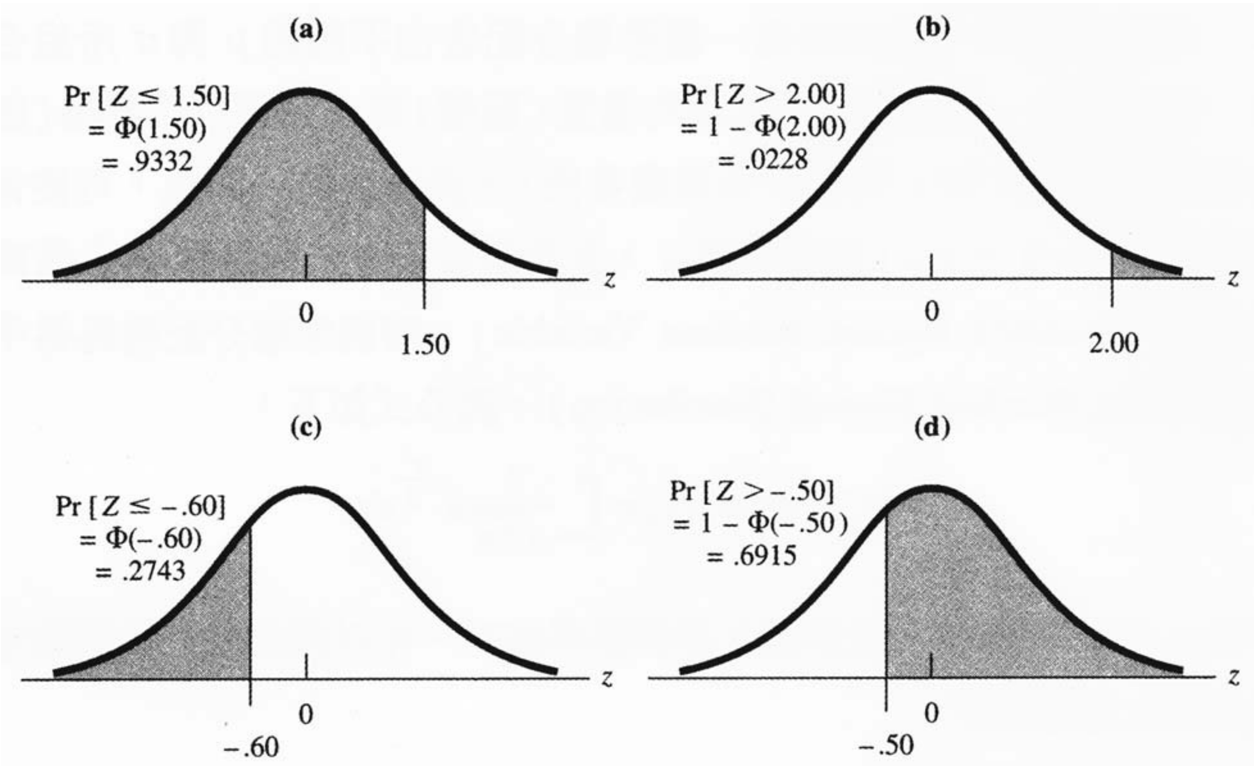
以平均數 μ 決定曲線的中心位置，而標準差 σ 則用以衡量曲線的散佈程度。

圖 7-9　三種不同平均數與標準差之母體常態曲線

◎標準常態分配

可使常態轉換成 $\mu=0$ 及 $\sigma=1$ 的標準值**Z**，此時常態變數**Z**稱為**標準常態隨機變數**(Standard Normal Random Variable)，轉換常態分配稱為**標準常態分配**(Standard Normal Distribution)，

$$\Phi(z)=\Pr[Z\le z]=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$



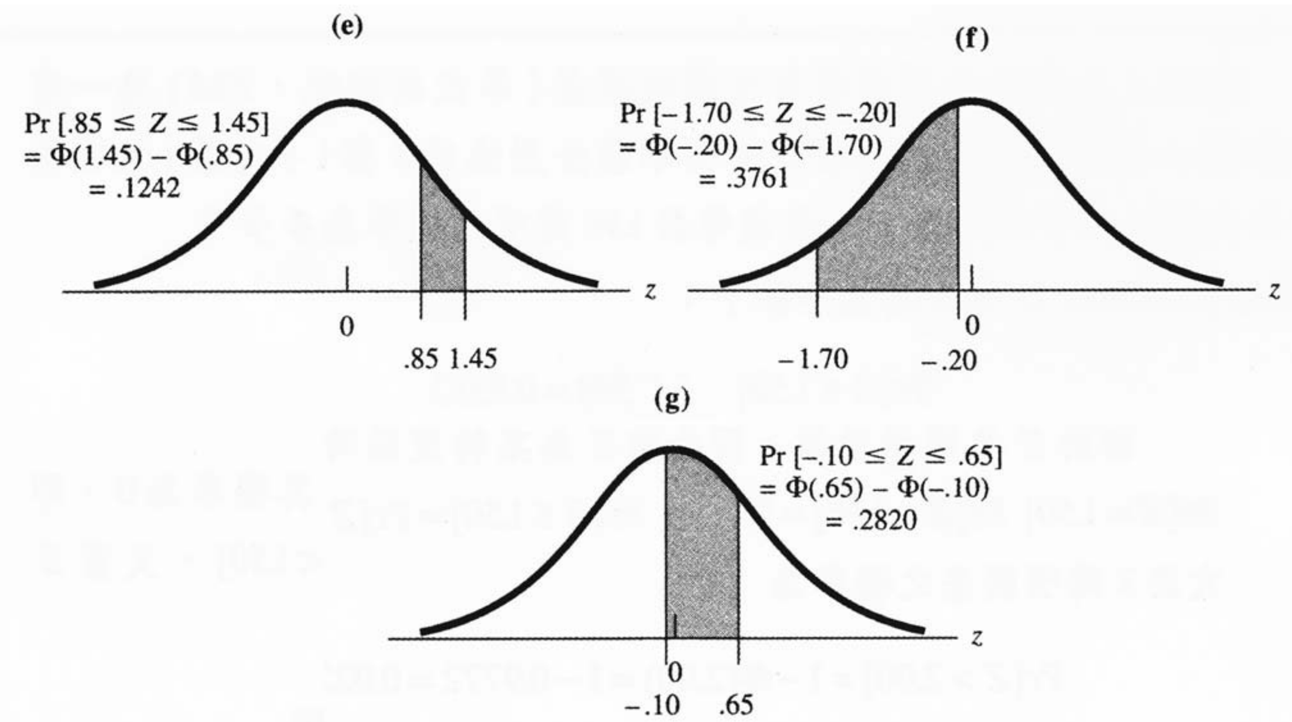


圖 7-11 不同的標準常態隨機變數範圍所對應之面積(機率)

◎任一常態隨機變數的機率

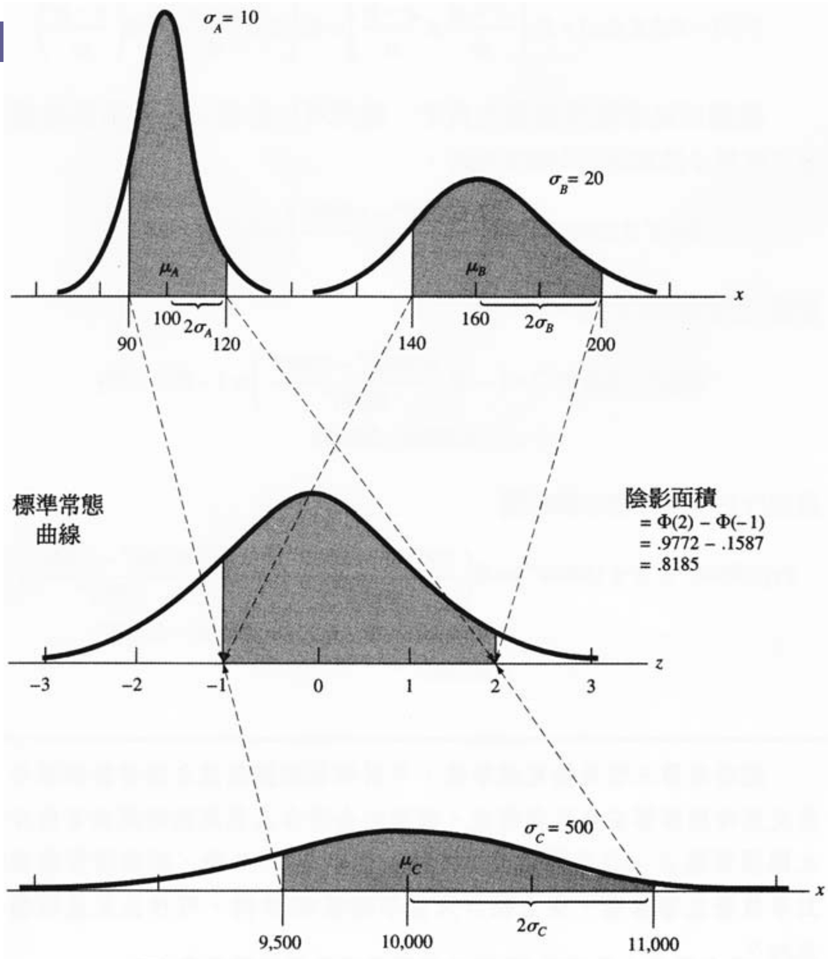


圖 7-12 三個不同 μ 及 σ 的常態分配的標準化過程

4-3:隨機抽樣

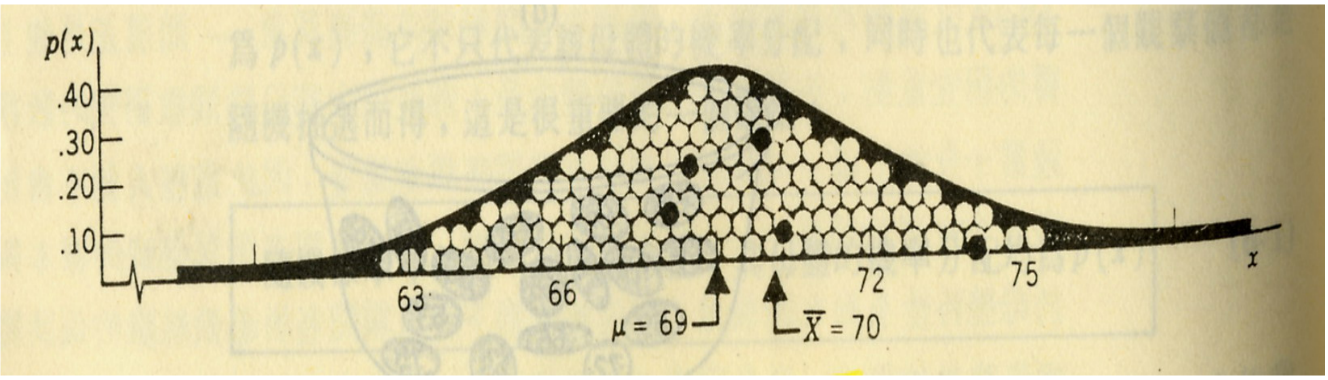
隨機樣本

非常簡單的隨機樣本是指一樣本中n個觀察值 $X_1, X_2, \cdots, X_n$ 彼此均相互獨立，此時，每一 X 的分配即為母體之分配 $p(x)$，亦即：

$$p(x_1)=p(x_2)=\cdots=p(x_n)$$

$$=\text{母體分配：}p(x)$$

因此每一個觀察值都有一母體平均數 μ 與標準差 σ。



自一母體抽出一典型的隨機樣本，此一樣本之平均數將個別的觀察值予以平均化－非極端化。

樣本可靠嗎？

當然，抽樣的目的是對於現行的母體進行統計推論。例如，我們總是希望樣本平均數 $\overline{X}$ 盡可能地接近母體平均數 μ。關於這一點，我們可以下列兩個方式來研究 $\overline{X}$ 如何趨近於 μ。

- 方法一：自母體抽出一樣本，然後計算其平均值，重複做這樣過程，並不斷計算出新的平均值X，最後觀察每一樣本X之變動情形，此即為X的抽樣分配。

- 方法二：用數學推導來求得X的抽樣分配。

樣本平均數的動差

樣本平均數$\overline{X}$之定義為：

$$\overline{X} \equiv \frac{1}{n}\left[X_1 + X_2 + \cdots + X_n\right]$$

由於$\overline{X}$是 $X_1, X_2, \cdots, X_n$ 的線性組合，

故$\overline{X}$本身亦為隨機變數。

！！那麼它的變動情形為何？

！！期望值與變異數又各為若干呢？

4-4 母體為常態分配時 $\overline{X}$ 的抽樣分配

$\overline{X}$ 之期望值為：

$$E(\overline{X}) = \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right]$$

每一觀察值X之母體分配為*p(x)*，而母體平均數為 μ，故 $E(X_1) = E(X_2) = \cdots = \mu$，因此：

$$E(\overline{X}) = \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right]$$

$$= \frac{1}{n}\left[n\mu\right] = \mu$$

$$\boxed{E(\overline{X}) = \mu}$$

$$\operatorname{var} \overline{X} = \frac{1}{n^2}\left[\operatorname{var}(X_1) + \operatorname{var}(X_2) + \cdots + \operatorname{var}(X_n)\right]$$

由於每一觀察值 *X* 存在一母體分配 *p(x)*，其中變異數為 σ^2，故：

$$\operatorname{var} \overline{X} = \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \cdots + \sigma^2\right]$$

$$= \left[n\sigma^2\right] = \frac{\sigma^2}{n}$$

$\overline{X}$ 的標準差 $= \dfrac{\sigma}{\sqrt{n}}$

此一 $\bar{X}$ 與 μ之離差乃表示估計誤差，即通常所稱之標準誤(standard error)或SE：

標準誤 $\overline{X}$ $$SE = \frac{\sigma}{\sqrt{n}}$$

母體平均值＝ 抽樣平均值＋抽樣誤差

平均數的範圍

1. 以 $\sigma_{\overline{X}}$ 代表樣本平均值之標準誤（Standard Error of the Sample Mean），此值與 σ 及樣本數有關。
2. 根據常態分布法則，可知約有95%樣本平均值落於 $\mu \pm 2\sigma_{\overline{X}}$ 管制界限內（即母體平均值的正負2個標準誤），以及約有99%之樣本平均值落於$\mu \pm 3\sigma_{\overline{X}}$ 管制界限內。

3. 若樣本平均重量落於管制下限（Lower Control Limit）LCL及管制上限（Upper Control Limit）UCL內，則LCL≦ $\overline{X}$ ≦UCL，稱製程在管制狀態中；反之，稱製程不在管制狀態中。

◎$\overline{X}$的常態分配

當自(μ,σ)常態分配之母體中抽樣，則$\overline{X}$的抽樣分配亦為常態分配，此時$\overline{X}$的平均數等於母體的μ。

如果母體σ未知，則以樣本s取代，亦不影響其性質，此時$\overline{X}$之標準誤為

$$\sigma_{\overline{X}} = \sigma / \sqrt{n}$$

◎**標準誤的特性**

(1) $\sigma_{\bar{X}}$ 是樣本 $\overline{X}$ 與母體 μ 間離散程度之測度，如果離散程度減少，此乃意謂著當 $\sigma_{\bar{X}}$ 減少時，任何單一樣本 $\overline{X}$ 將更接近母體 μ。

(2) 當母體為常態分配時，$\sigma_{\bar{X}}$ 值取決於 σ 與樣本 n 數。由於 $\sigma_{\bar{X}} = \sigma / \sqrt{n}$，故當母體 σ 愈大，則 $\sigma_{\bar{X}}$ 亦愈大；反之，若 σ 愈小，則 $\sigma_{\bar{X}}$ 愈小。

(3) 第二個會影響 $\sigma_{\bar{X}}$ 值者為樣本大小。

4-5 非常態母體的抽樣分配

◎若當母體並非為常態分配，只要樣本數夠大（通常指n>30），依據**中央極限定理（Central Limit Theorem）**，$\overline{X}$ 的抽樣分配會近似於常態分配。

◎**中央極限定理：**

所謂中央極限定理係指從一母體中隨機抽樣，不限母體的分配型態為何，只要樣本數夠大（一般為n>30），則不論母體的機率分配型態，樣本平均數 $\overline{X}$ 的抽樣分配近似於常態分配。

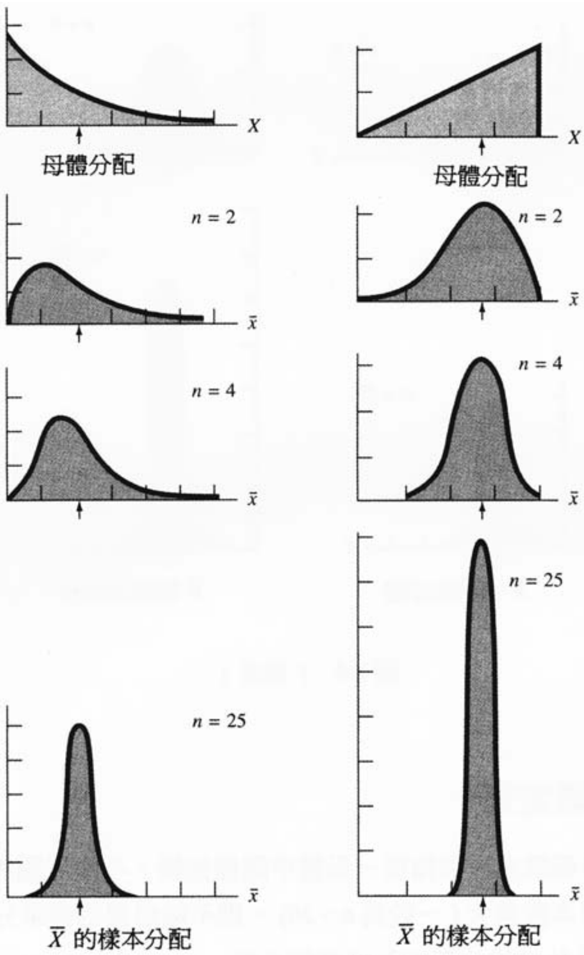


圖 9-4　中央極限定理的說明（不同 n 之 $\bar{X}$ 的抽樣分配—中央極限定理）

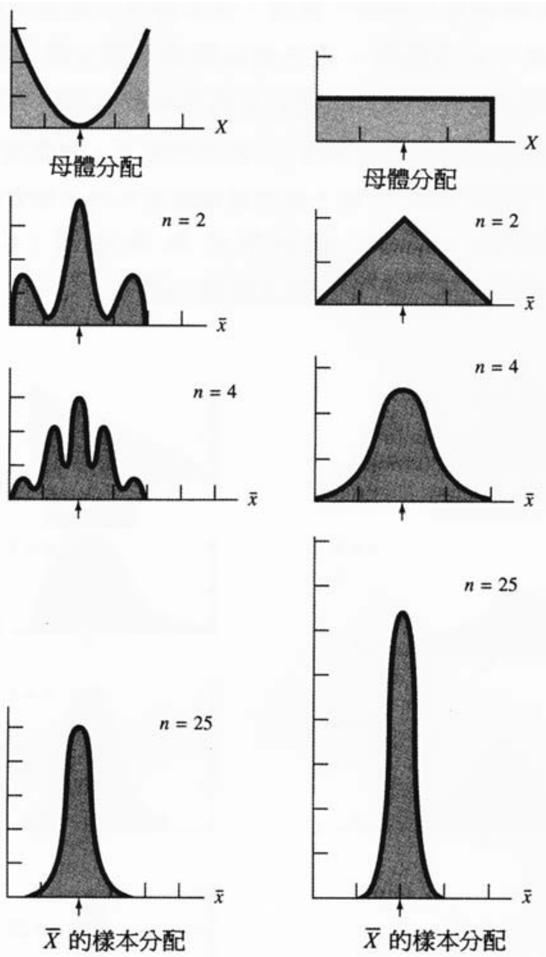


圖 9-4　（繼續）

◎當母體σ未知時

在這種情形下，我們可

(1)以與此母體特性相近的另一已知母體 σ 來代表。

(2)在大樣本(n>30)條件下，無論母體分配是否屬於常態，則以樣本標準差 S 取代之。

(3)當母體為常態分配及小樣本情形下，可用樣本S取代，惟此時須用Student-t 分配來分析。

◎ t 分配的統計量

(1)由於母體參數通常為未知，故其參數值另以樣本統計量來估計：即用$\overline{X}$估計μ，以 s 估計 σ。

(2)當以樣本 s 取代σ，則樣本$\overline{X}$之標準常態分配之轉換式為

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

由於$\overline{X}$與 s 皆為隨機變數，故經轉換後$\overline{X}$的抽樣分配會隨著樣本數的不同而改變，故上式亦為一隨機變數。

(3)在小樣本情形下，以 s 估計 σ 並不很精確。針對此問題，統計學家W.S. Gosset 在1908年以student為筆名的論文中首先提出 t 分配，或簡稱 t 分配。 t 分配的統計量(Student Statistic)為

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

(4)圖9-5的 t 分配曲線中，α 面積即為在 t_α 點(值)以右(右尾)的面積，其面積代表其機率值，可表為下式：

$$\alpha = \Pr[t > t_\alpha]$$

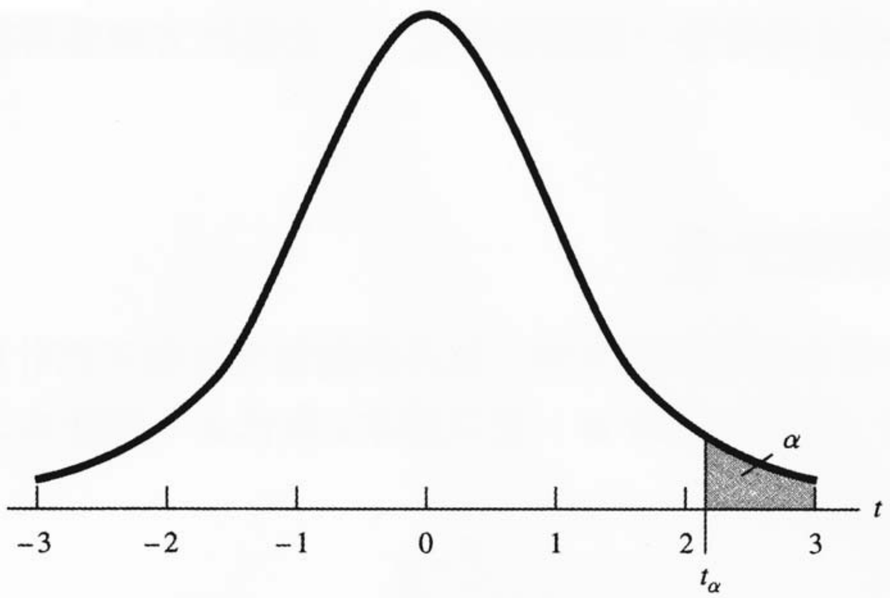
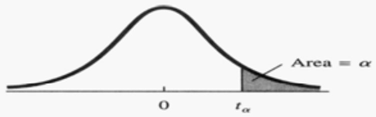


圖 9-5　t 分配之機率密度曲線 (df = 4)

Table G
Student *t* Distribution



The following table provides the values of t_α that correspond to a given upper-tail area α and a specified number of degrees of freedom.

Degrees of Freedom	Upper-Tail Area α .4	.25	.1	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1 310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Source: E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*. Vol. I. London: Cambridge University Press, 1966. Partly derived from Table III of Fisher and Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Longman Group Ltd., London (previously published by Oliver & Boyd, Edinbugh, 1963). Reproduced with permission of the authors and publishers.

◎ *t* 分配曲線與常態曲線之比較

*t*分配與常態分配相似，兩者都是對稱於0的平均值之鐘形分配。

(1) *t* 統計量受到兩個隨機變數 $\overline{X}$ 與 s^2 的影響，而Z統計量只隨 $\overline{X}$ 的變動而改變。

(2) Z分配的母體 σ 為已知而 *t* 分配時 σ 未知，故*t*分配的變異性較大。

(3)圖9-6可看出兩者皆為相似的鐘形曲線，但*t*分配的曲線的頂端較Z分配為低，且*t*分配曲線的尾端較Z分配為分散，表示其變異性較大。

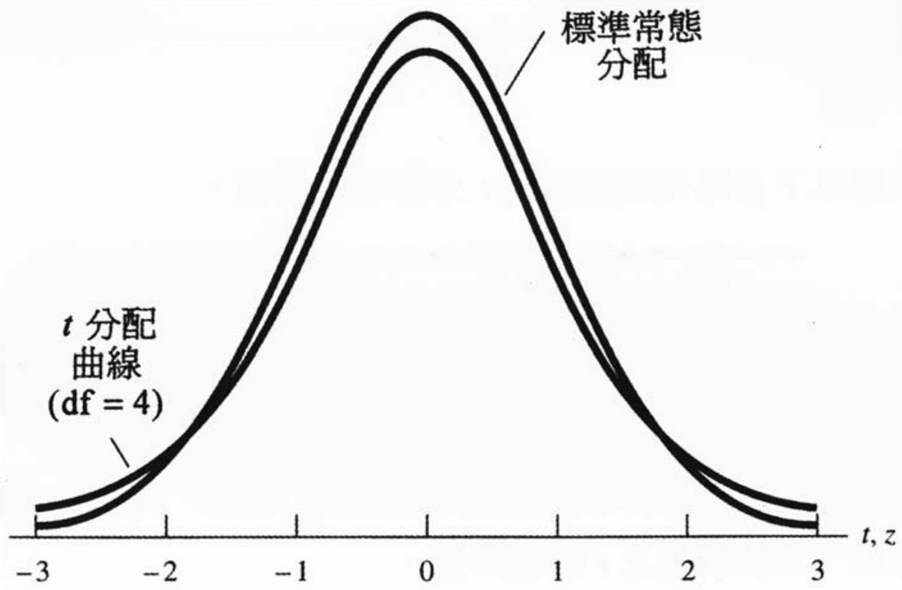


圖 9-6　標準常態密度曲線與 t 分配曲線之比較