偶然誤差理論

機率理論以及各種分配的特性
標準誤差的機率

概述

- 測量平差的主要任務是，對僅含偶然誤差的觀測量，如何進行平差
  - 在測量中觀測量中的系統誤差可於作業前或作業後進行改正，同時測量作業要想辦法排除錯的觀測量，因此觀測量中一般均視爲僅含偶然誤差
- 本章主要在推導偶然誤差的一般理論，以及簡單的錯誤排除方法
  - 詳細的錯誤排除方法內容將於第20章敘述

機率理論

- 機率
  - 事件發生的機會，一般以事件發生的次數除以事件所有會發生的總數來表示
    - 如擲骰子出現二點的機率為1/6，因為骰子二點只有一面，而骰子總共有六面
    - 又如某事件的出現有*m*種方式，不會出現有*n*種方式，則其出現之機率為$m/(m+n)$，不會出現之機率為$n/(m+n)$
  - 機率的值介於0與1之間，0表示不可能發生，1則爲一定會發生
  - 複合事件爲兩個或多個同時出現之事件
    - 在測量中常發生複合事件，如角度與距離的偶然誤差導致導線的閉合差
    - 複合事件的機率為各個單獨事件發生機率的乘積

機率理論

- 例如， *A*盒有四個球，一紅三白；*B*盒有五個球，二紅三白。自兩盒隨機各取一球，爲兩紅球之機率有若干？
  - 自A盒取出紅球的機率為1/4，自B盒取出紅球的機率為2/5，因此取出兩紅球的機率為$1/4 \times 2/5 = 2/20$
  - 同理，同時各取一球為兩白球之機率是：$3/4 \times 3/5$，或$9/20$；為一白一紅之機率是：$1 - (2/20 + 9/20) = 9/20$
- 引伸至任意個數事件$P = P_1 \times P_2 \times \ldots \times P_n$

二個紅球的機率　　二個白球的機率

機率理論

- 偶然誤差發生的原理
  - 以捲尺丈量A、B兩點間的距離爲例
    - 假設丈量僅含大小為1的偶然誤差
    - 每次丈量的誤差不是+1就是-1，因此誤差出現+1或-1的機率各為1/2
    - 若往返丈量，則誤差出現的可能組合為： -1與-1 ，-1與+1 ，+1與-1 ，+1與+1， 而最後的誤差分別為-2, 0, +2，因此最後結果的誤差為-2的機率為1/4， 為0的機率為1/2， 為+2的機率為1/4
    - 若丈量的次數為n，則誤差的可能組合將為2^n
      - 如丈量3次，則可能組合數爲$2^3=8$；丈量4次，則可能組合數爲$2^4=16$等等

機率理論

- 表3.1為根據上述分析的結果，圖3.1則為其對應直方圖
  - 橫軸爲最後結果的誤差值，縱軸爲其機率
- 若n逐漸增大，則直方圖將逐漸地趨向一鐘形的平滑曲線

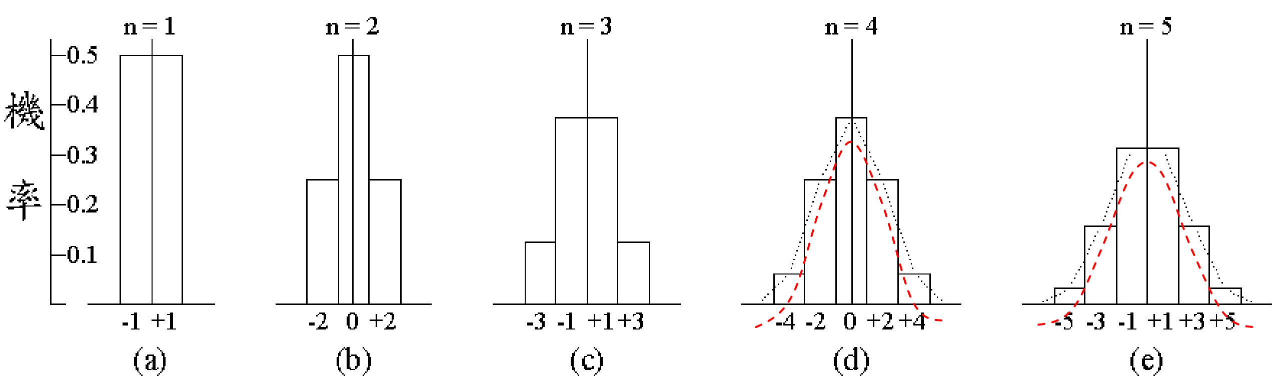

圖3.1 機率與成果誤差大小之對應

表3.1 偶然誤差事件

(1) 量測組數	(2) 成果誤差	(3) 頻率, t	(4) 所有個數, T	(5) 機率
1	+1	1	2	1/2
	-1	1		1/2
2	+2	1	4	1/4
	0	2		1/2
	-2	1		1/4
3	+3	1	8	1/8
	+1	3		3/8
	-1	3		3/8
	-3	1		1/8
4	+4	1	16	1/16
	+2	4		1/4
	0	6		3/8
	-2	4		1/4
	-4	1		1/16
5	+5	1	32	1/32
	+3	5		5/32
	+1	10		5/16
	-1	10		5/16
	-3	5		5/32
	-5	1		1/32

機率理論

- 此鐘形平滑曲線稱為常態分配曲線，又稱為常態隨機變數之機率密度函數
  - 曲線下面積等於1，代表所有誤差發生之機率總和
  - 常態機率密度函數，如式*(3.2)*所示，詳細的推導過程見本書附錄D的D.1節

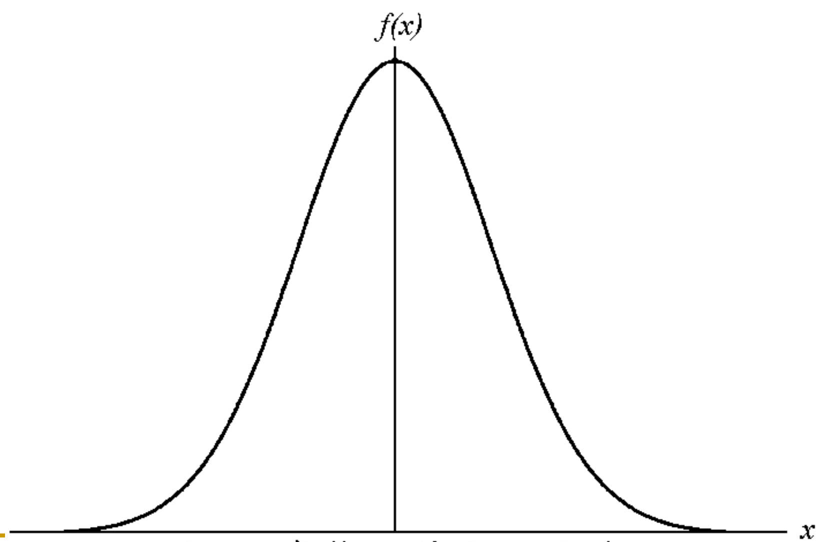

圖3.2 常態誤差分配曲線

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} \qquad (3.2)$$

常態分配曲線的特性

- 在(3.2)式中， *f*(*x*)為x大小在*x*與*x*+*dx*之間，誤差出現的機率
  - 此誤差機率相當於在*x*與*x*+*dx*之間，曲線下與橫軸所爲的面積
  - 而曲線下全部的面積爲1（見(3.3)式）
- 爲了了解此曲線的特性，必須由此曲線函數的微分著手

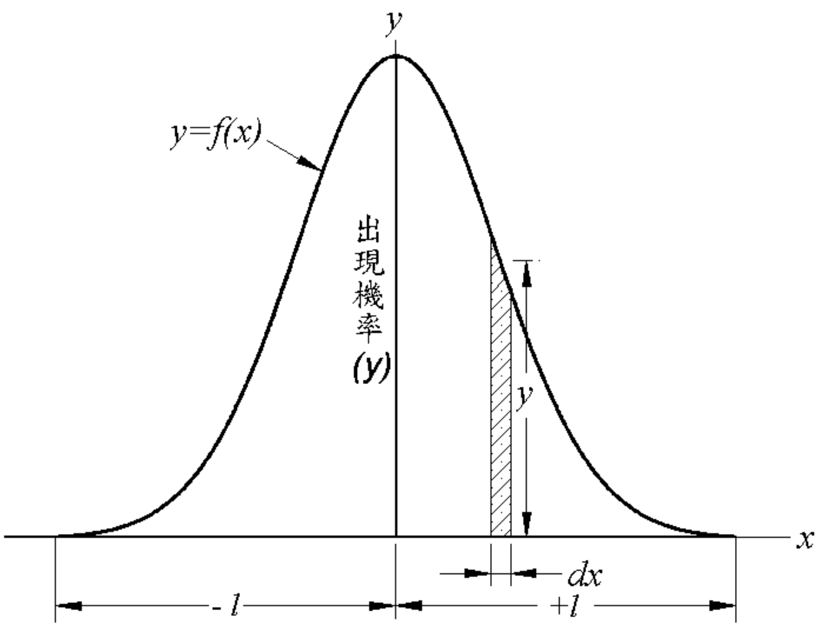

圖3.3 常態密度函數

$$area = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} = 1 \qquad (3.3)$$

常態分配曲線的特性

- 令y=f(x)，則其一次微分如(3.5)式所示
- 而其二次微分如(3.8)式所示
- 當一次微分為0時，表示曲線的極值出現
  - 由(3.5)式，當x=0或y=0時，出現極值
  - 表示曲線在中央處其切線平行於x軸
- 當二次微分為0時，曲線的斜率產生轉折
  - 由(3.8)式，當 $x^2/\sigma^2 - 1=0$時，曲線產生轉折，即$x=\pm\sigma$處爲曲線之轉折點
- 當x=0時，y的值如(3.9)式所示
  - 此爲曲線中央處的縱坐標，由式中可看出y與σ成反比

$$\frac{dy}{dx} = -\frac{x}{\sigma^2}\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right) \quad (3.4)$$

$$\frac{dy}{dx} = -\frac{x}{\sigma^2}y \quad (3.5)$$

$$\frac{d^2y}{dx^2} = -\frac{x}{\sigma^2}\frac{dy}{dx} - \frac{y}{\sigma^2} \quad (3.6)$$

$$\frac{d^2y}{dx^2} = \frac{x^2}{\sigma^4}y - \frac{y}{\sigma^2} \quad (3.7)$$

$$\frac{d^2y}{dx^2} = \frac{y}{\sigma^2}\left(\frac{x^2}{\sigma^2} - 1\right) \quad (3.8)$$

$$y = \frac{1}{\sigma\sqrt{2\pi}} \quad (3.9)$$

標準常態分配函數

- 對常態分布的密度函數積分可獲得其機率，即分配函數，如(3.10)式所示
  - 式中t為積分時的上限
  - 為了方便求得積分結果，可事先計算並做成表
  - 然此積分的結果受到μ與σ^2影響，不利於製表
  - 若令$\mu=0$，變方$\sigma^2=1$，則可方便製表，如本書附錄D.1所示，以利後續之使用
  - 此種$\mu=0$，$\sigma^2=1$的常態分配稱之為標準常態分配
  - 附表D.1中最左欄位之t值即圖3.4中所示，t之單位為σ，表中最上一列為t值的小數點第二位值
    - 表列值為由-∞至t時，在標準常態曲線下之面積，譬如求由-∞至1.68曲線下之面積為0.95352

$$F_x(t)=\int_{-\infty}^{t}\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}dx \qquad (3.10)$$

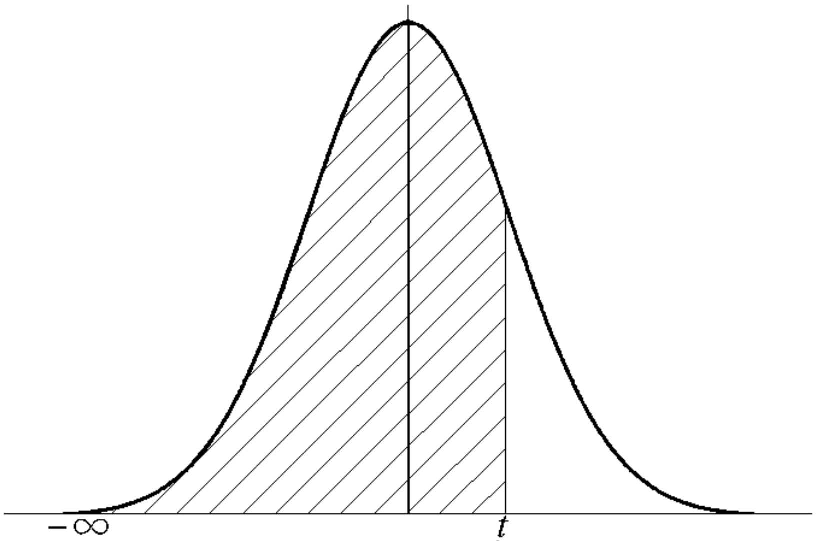

圖3.4 利用式(3.10)求得之面積

標準常態分配函數

- 表D.1可應用於任何的μ與σ^2
  - 如y爲N (μ,σ^2)，並令$z=(y-\mu)/\sigma$，則z爲N(0,1)，即標準常態分配
  - 將z變數的$\mu=0$、$\sigma^2=1$代入(3.2)式，得(3.11)式，而其標準常態分配函數則爲(3.12)式

$$N_z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad (3.11)$$

$$N_z(z) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \qquad (3.12)$$

標準常態分配函數

- 假設z為常態隨機變數，則$z<t$之機率為：$P(z<t)=N_z(t)$
  - 如圖3.5所示，在求算介於a與b之間t的面積(機率)時，可分別計算$z<a$與$z<b$之機率[即$P(z<a)=N_z(a)$與$P(z<b)=N_z(b)$]，兩者之差異即斜線部分的面積：$P(a<z<b)=N_z(b) – N_z(a)$
    - 若兩界限值大小相同，符號相反時(此即$-a=b=t$)，則機率為：$P(|z|<t)=N_z(t) – N_z(-t)$

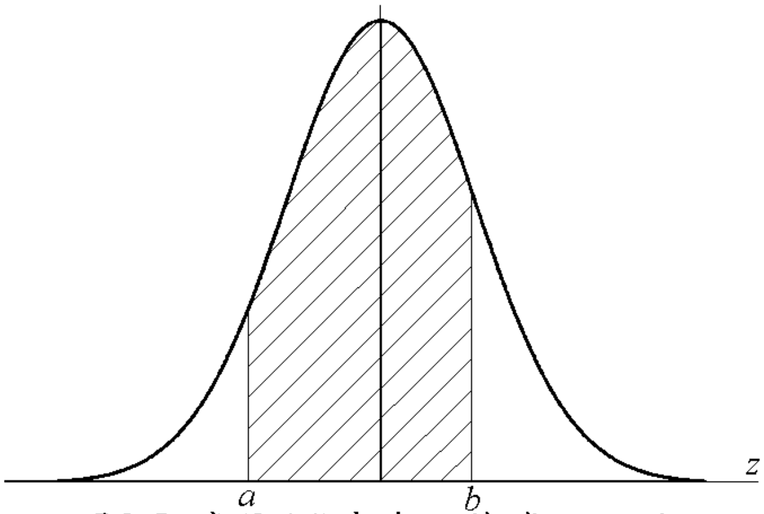

圖3.5 式(3.14)中表示機率之面積

標準常態分配函數

- 因為常態分配的對稱性，所以存在$P(z>t)=P(z<-t)$ 的關係，如圖3.6所示
  - 如 $t=-1.00$之表列值爲：0.15866，$t=+1.00$之表列值則爲：0.84134，而因最大機率(面積)爲1，故大於$+1.00$之面積爲$1-0.84134= 0.15866$，此與小於-1.00之面積相同
  - 由此可獲得：$1 - N_z(t)=N_z(-t)$
  - 因此得知： $P(|z|<t) = 2N_z(t) - 1$

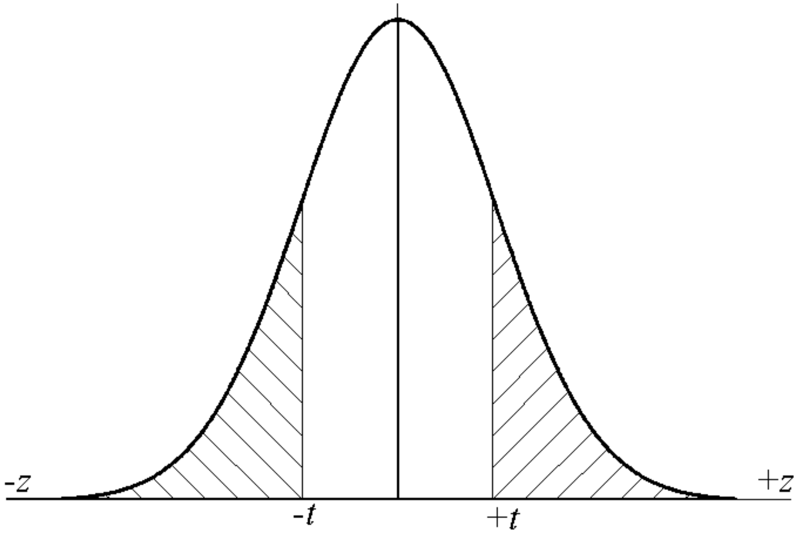

圖3.6 式(3.16)中表示機率之面積

標準誤差的機率

- 標準誤差的機率，是指在±σ界線範圍內的機率
- 對標準常態分配而言，$\sigma^2=1$；查$t=-1(\sigma=-1)$與$t=1(\sigma=1)$於附表D.1之表列值分別為：0.15866與0.84134
  - $P(-\sigma<z<+\sigma)=N_z(+\sigma)-N_z(-\sigma)=0.84134-0.15866=0.68268$
  - 所有觀測量有68.3％之可能落在-σ與+σ之間
  - 對一組觀測量而言，任一觀測的誤差，有68.3％之可能在-σ與+σ之間
  - 上述的說明，可應用於任何具有常態分配的觀測量集合。
  - 由於常態分配曲線的轉折點是在±σ處，這一點也可由圖3.7看出

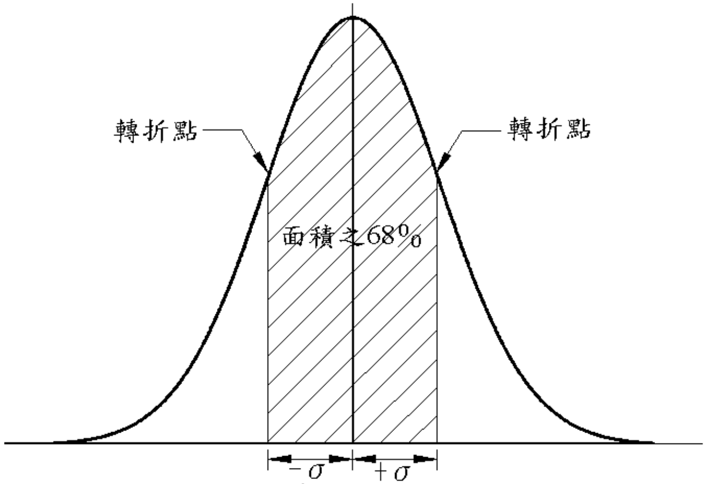

圖3.7 常態分配曲線

50%或是誤差

- 對一組觀測量而言，50%或是誤差是在建立落於誤差50%的界限
  - 換言之，任何量測在此範圍內的機會與範圍外的機會相等
- 因*50%*或是誤差的機率為*1/2*，由(3.18)是得$P(|z|<t) = 2N_z(t) - 1 = 0.5$ 或$N_z(t) = 0.75$
  - 查D.1表得$N_z(0.67) = 0.7486$ 且 $N_z(0.68) = 0.7517$；利用內插方式可反求得*t*值
    - $\Delta t/(0.68-0.67)=(0.75-0.7486)/(0.7517-0.7486)$
    - $\Delta t=0.0045$
    - $t= 0.67+0.0045=0.6745$
- 對任意觀測組，*50%*或是誤差可將標準偏差乘以所求之*t*值而求得
  - $E_{50}=0.6745\sigma$

95%或是誤差

- *95％*或是誤差(E_{95})，理論上是在界定*95％*的誤差的範圍
  - 常被測量人員用來說明觀測量的精密度
  - 與50%或是誤差求界限的方式相似
    - $P(|z|<t) = 2N_Z(t) - 1 = 0.95$ 或 $N_Z(t) = 0.975$
    - 查表並利用內插方式，可得$t=1.960$
    - 對任意觀測組，*95％*或是誤差為$E_{95}=1.960\sigma$

其他百分比的或是誤差

- 利用類似的計算過程，可得其他百分比的或是誤差
  - 常用來偵錯之$E_{99.7}=2.965\sigma$
  - 常用的百分比或是誤差
    - $E_{50}=0.6745\sigma$
    - $E_{90}=1.6449\sigma$
    - $E_{95}=1.960\sigma$
    - $E_{99}=2.576\sigma$
    - $E_{99.7}=2.965\sigma$
    - $E_{99.9}=3.29\sigma$

百分比誤差的用途

- 評估測量結果是否被接受
  - 常使用的有*90％*誤差或*95％*誤差
    - *95％*誤差，又稱兩倍標準差(*two-sigma*，*2σ*)，實際上是*1.960σ*
- 觀測量的偵錯
  - 一般超出$\mu \pm 3\sigma$範圍之觀測量可視爲錯誤而去除之

例子

- *例3.1* 若有15個獨立量測之距離值(D_i，單位為*m*)：*64.684, 64.693, 64.687, 64.663, 64.657, 64.651, 64.705, 64.721, 64.684, 64.690, 64.675, 64.693, 64.699, 64.678* 與*64.693*，計算平均值、S、E_{50}、E_{95}，並檢核是否有任何量測超過*99.7%*之水準

- *解：*

$$\bar{y} = \frac{\sum D_i}{n} = \frac{970.273}{15} = 64.685m$$

$$S = \sqrt{\frac{62761.9843 - 15 \times 64.685^2}{15 - 1}} = \pm 0.018m$$

檢視這些資料，有*10*個量測位於*64.685±0.018*或*(64.667, 64.703)*之間，對應有*10/15×100*或*66.7%*之量測個數，若符合常態誤差分配理論，這是可預期的結果

由式(3.20)，

$E_{95}=1.960\ S=\pm 1.960\times 0.018=\pm 0.036m$；所有資料均已位於64.685±0.036或(64.649, 64.721)之間。

在99.7％之信心水準下，±2.965S對應之區間範圍為：±0.054，所有數值都已在這個範圍內，因此沒有理由相信有任何量測為大錯甚至更大的錯誤

根據式(3.19)，

$E_{50}=0.6745\ S=\pm 0.6745\times 0.018=\pm 0.012m$

再檢視這些資料，有9個量測位於64.685±0.012或(64.673, 64.697)之間，對應有9/15×100或60%之量測個數，比常態分配理論應有之50％略高；惟須注意這些資料僅為母數群中之樣本，不得即以此作為拒絕整體資料集合的理由(第四章將討論樣本集合之統計區間等)

*例*3.2 若利用1"讀儀器觀測方向，得下列50個秒位數，計算平均值、標準差與E_{95}，並檢核是否有任何超過99%之觀測錯誤。

41.9 46.3 44.6 46.1 42.5 45.9 45.0 42.0 47.5 43.2 43.0 45.7 47.6 49.5 45.5 43.3 42.6 44.3 46.1 45.6 52.0 45.5 43.4 42.2 44.3 44.1 42.6 47.2 44.7 44.2 46.3 49.5 46.0 44.3 42.8 47.1 44.7 44.7 45.6 45.5 43.4 45.5 43.1 46.1 43.6 41.8 44.7 46.2 43.2 46.8

解：50個量測的和為：2252.0，故平均值為：

2252.0/50=45.04″，利用式(2.10)，標準差為：

$$S = \sqrt{\frac{101649.94 - 50 \times 45.04^2}{50-1}} = \pm 2.12''$$

上式中，Σy^2=101,649.94，有35個量測位於45.04±2.12或(42.92, 47.16)之間，對應為35/50×100=70%之量測個數，與預期之68.3%水準相當接近。

根據式(3.20)，$E_{95}=\pm1.960\times2.12=\pm4.15''$，資料中有3個量測位於$45.04\pm4.15$或(40.89, 49.19)之外，且都比49.19大，故有47/50或94%之量測個數位於E_{95}之範圍內。

如第16頁前述，$E_{99}=\pm2.576\times2.12=\pm5.46''$，資料中應有99%之量測個數位於$E_{99}$之範圍$45.04\pm5.46$或(39.58, 50.50)之內，實際上也僅有一個數值(52.0)落於此範圍之外，故有98%之量測個數位於此範圍之內。

由上述分析，可見資料集合向左偏移，此即：超過範圍的數值都在資料的右邊；由圖3.2可見資料之偏斜性，也因此，應將52.0一值視為錯誤，將之去除後，再重新計算得：

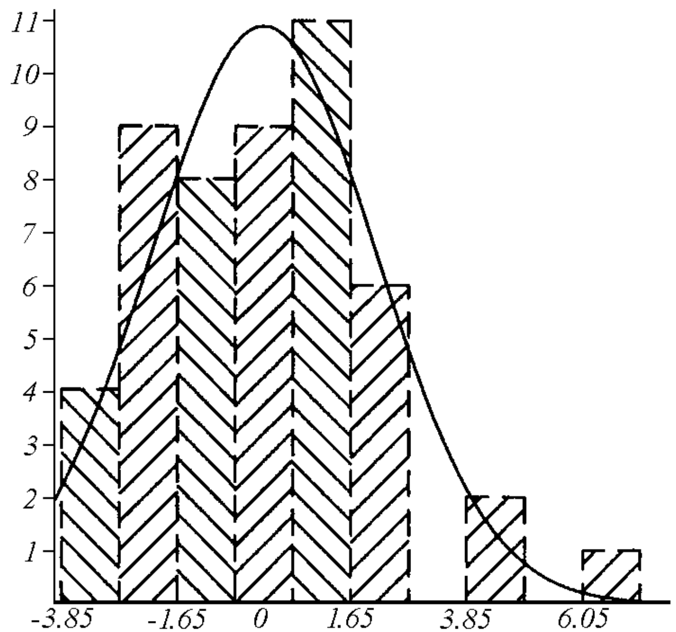

圖3.8 偏斜資料集

$$mean = \frac{2252 - 52}{49} = 44.90''$$

$$\sum y^2 = 101{,}649.94 - 52.0^2 = 98{,}945.94$$

$$S = \sqrt{\frac{98{,}945.94 - 49 \times 44.90^2}{49 - 1}} = \pm 1.88''$$

重新計算誤差後，有32個或量測個數之65.3%位於平均值$\pm S$範圍內；又有47個或量測個數之95.9%；位於平均值$\pm E_{95}$範圍內；此外，沒有任何數據在平均值$\pm E_{99}$之外。因此在99%信心水準下，沒有理由拒絕任一資料。