Chapter 4
Common Probability Distribution Functions

Introduction

- The application of probability distribution function is to *infer* the properties of population based on the sample analysis.
- The coverage factor k mentioned earlier in the topic of uncertainty can be 2 or 3. *Why?*
- Because it involves a probability distribution function called Normal probability distribution function which has a notation $N(\mu, \sigma)$.

Types of PDFs

- The probability distribution functions (PDFs) applicable to estimation of measurement uncertainty in chemical analysis are:
  - Normal (or Gaussian) probability distribution
  - Rectangular probability distribution
  - Triangular probability distribution
- Other common probability distribution functions are:
  - Poisson probability distribution
  - $\chi 2$ probability distribution

Normal Distribution Functions

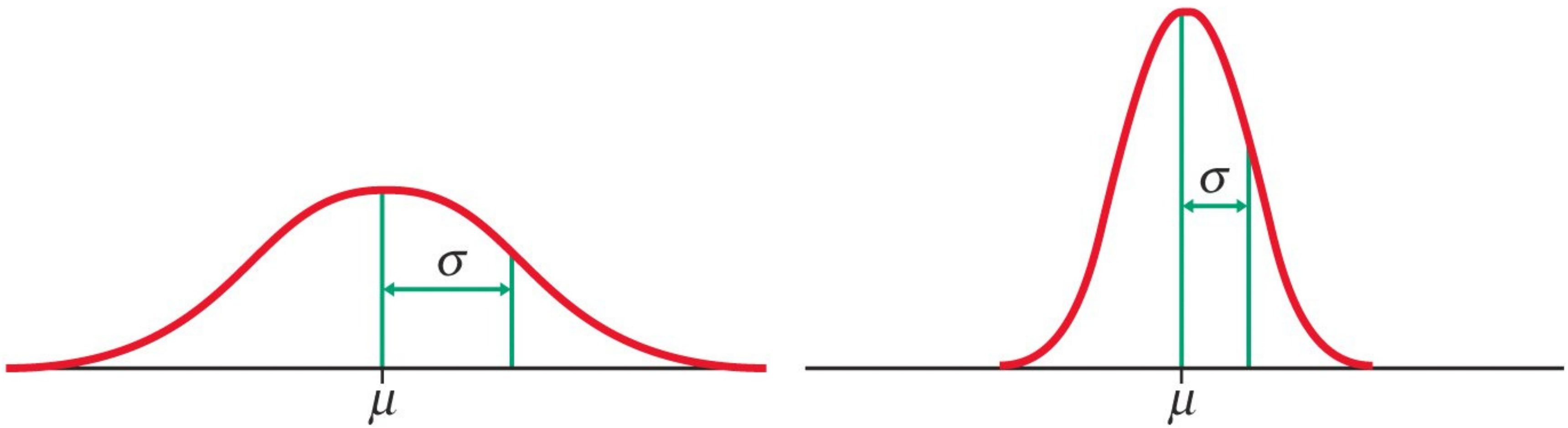
- The spread of random error of a chemical analysis falls in a Normal distribution
- **Normal or Gaussian Distribution**

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

where, y is the frequency density, μ is the population mean, σ, population standard deviation and e = 2.718.

The Normal Distribution

- Mean μ defines the center of the curve
- Standard deviation σ defines the spread
- Notation is $N(\mu,\sigma)$.



The Normal distribution function

- Irrespective of the values of μ and σ, the integrated area under the bell-shaped curve represents the total probability for all kinds of deviations (x_i-μ) being equals to 1 when -∞<x<+∞.

$$P(-\infty < x < \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

- Let

$$z = \frac{x - \mu}{\sigma}$$

The Normal distribution function

- By differentiation, we get $dz = \frac{dx}{\sigma}$ or $dx = \sigma.dz$
- Therefore,

$$f(x)dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \phi(z)dz$$

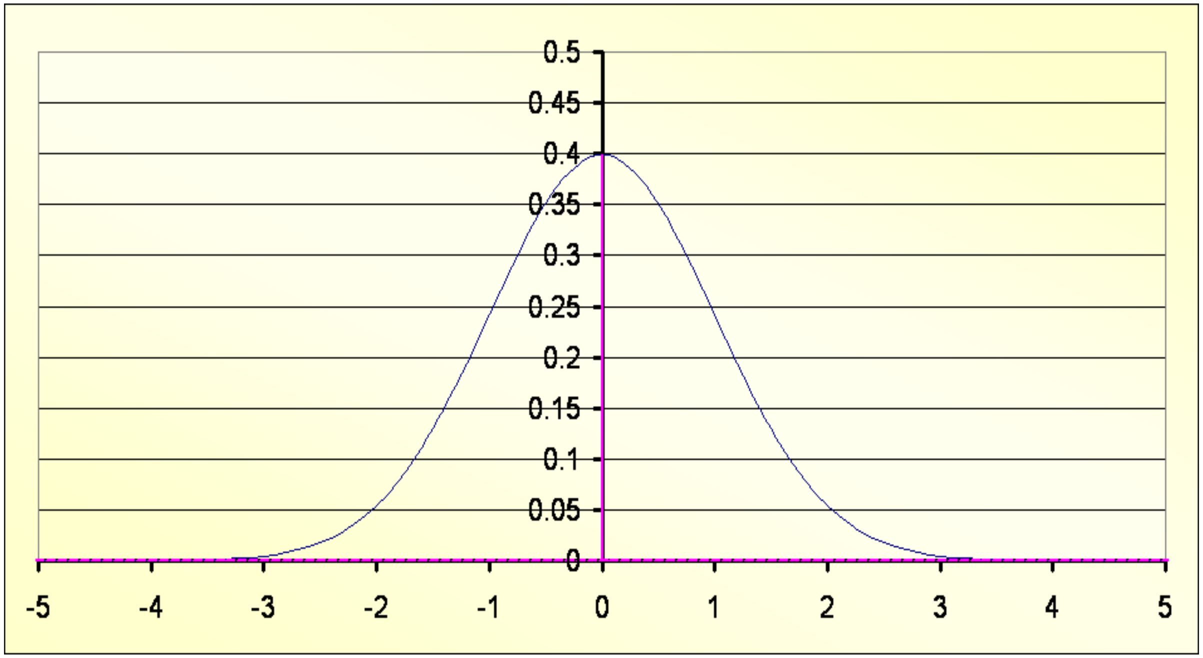
- and the new standardized function is :

$$y = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

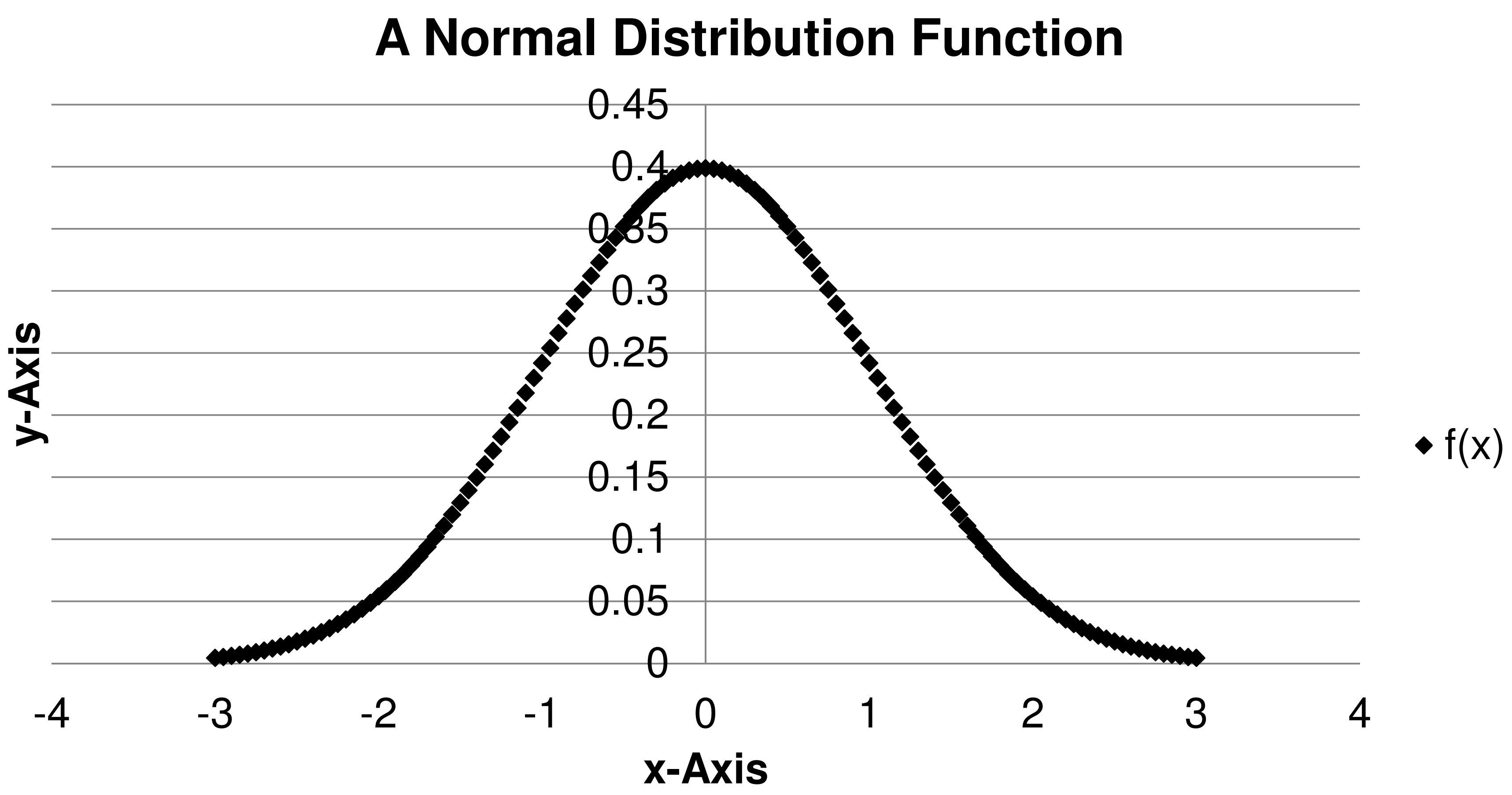
The Normal distribution function

- Here $N(\mu, \sigma)$ density parameter becomes $N(\mu = 0, \sigma = 1)$ or $N(0,1)$ with standardized

$$P(-\infty < x < \infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dx = 1$$



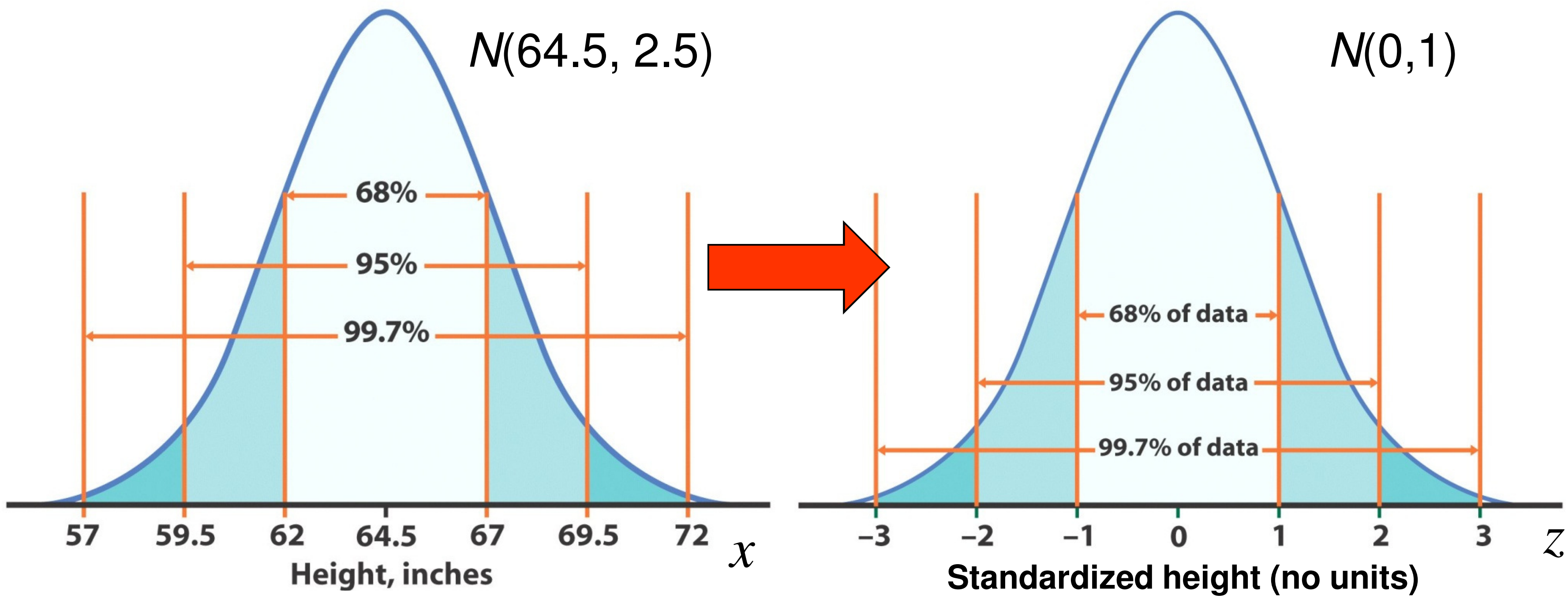
An Excel spreadsheet on Normal Distribution



Standardized *N* (μ,σ)

We **standardize** Normal data by calculating *z*-scores.

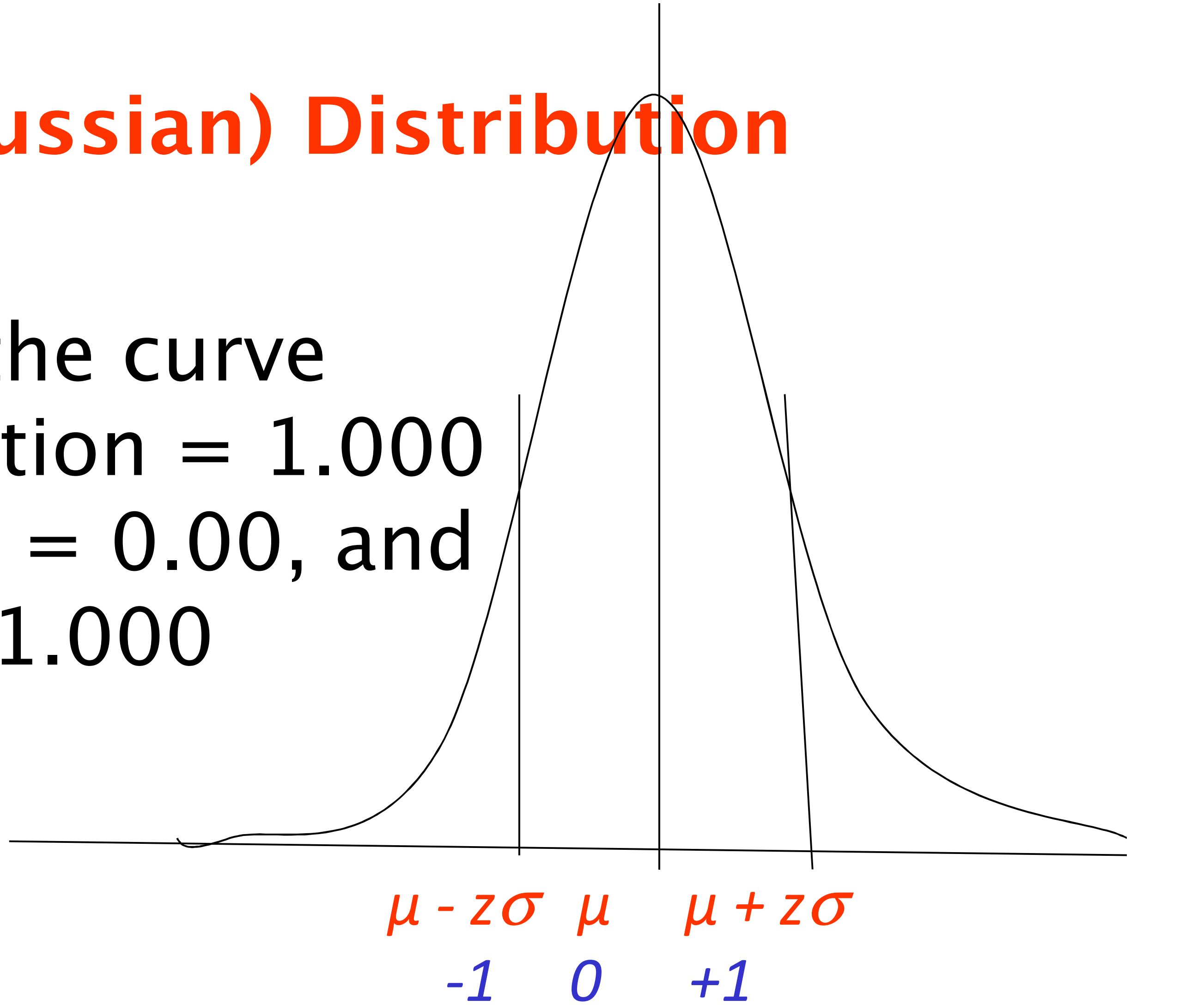
$$z = \frac{(x - \mu)}{\sigma}$$



Any N(μ,σ) can be standardized to a *N*(0,1).

The Normal distribution function

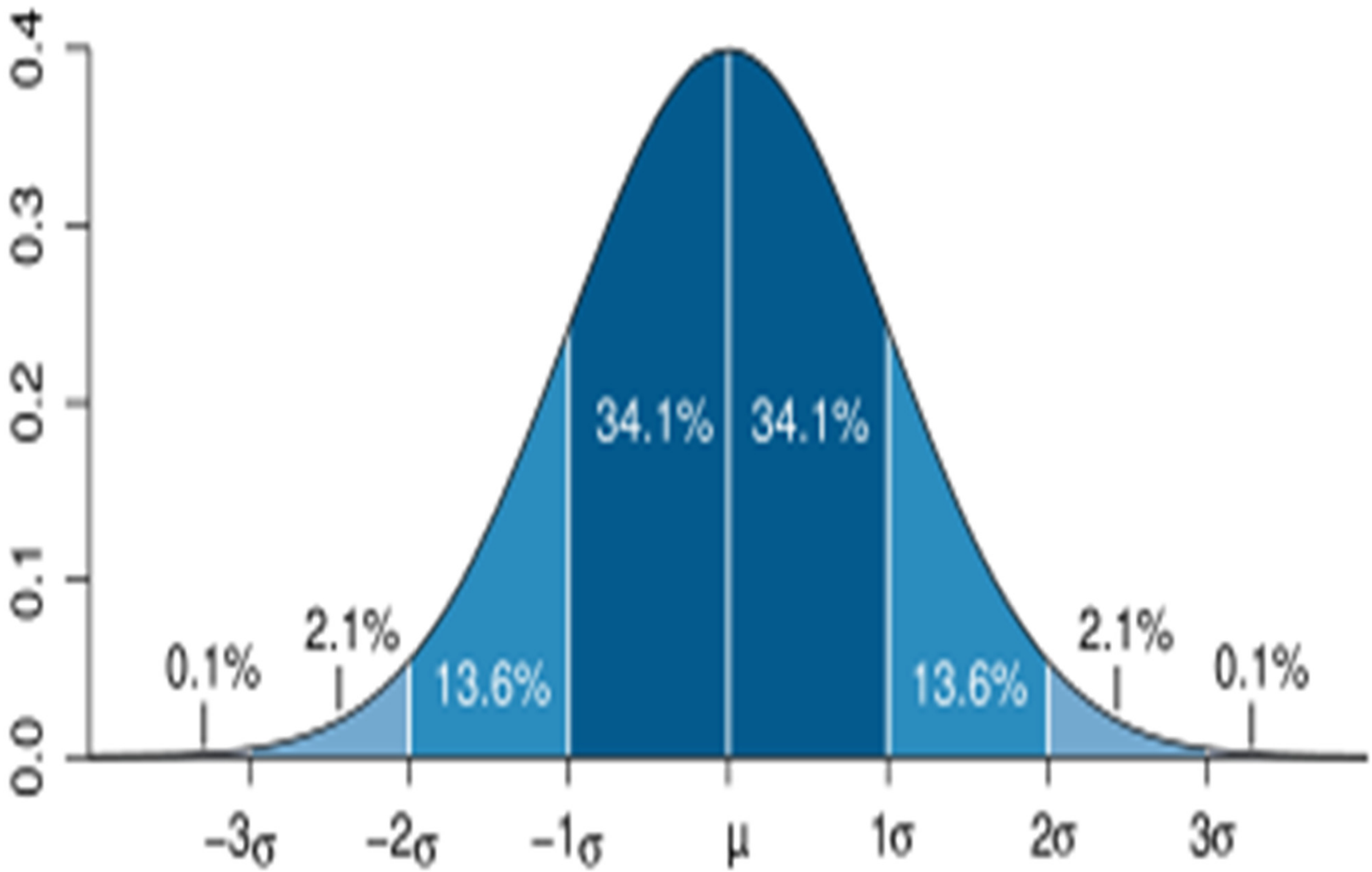
- **Normal (Gaussian) Distribution**
- *That means*:
- Area under the curve after integration = 1.000 with mean μ = 0.00, and std dev σ = 1.000



Probability Distribution Functions

- **Normal or Gaussian Distribution**
- It can also be shown that:
- Area between z = -1 and + 1: about 0.68
- Area between z = -2 and +2: about 0.95
- Area between z = -3 and +3: about 0.997
- Note: 0.95 area out of 1.000 means 95%
- To be exact, $-1.96 < z < +1.96$ to give an area of 95.00%
- *Learn how to use the Normal Distribution table*

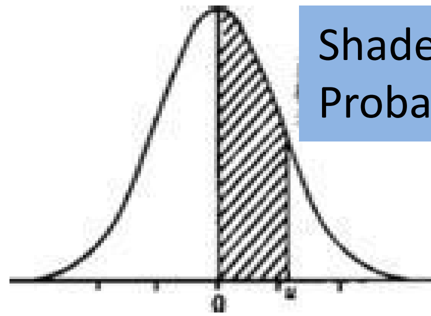
Obtaining Normal Probability in EXCEL:
To obtain: $F(x)=P(X\leq x)$, use Function: =NORMDIST(x,m,s,1)


0.4
0.3
0.2
0.1
0.0
34.1%
34.1%
13.6%
13.6%
2.1%
2.1%
0.1%
0.1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ

Example

- If $x < +2.35$,
- the EXCEL's function: *=NORMDIST(2.35,0,1,TRUE)* gives 0.991, *i.e*. the area under the curve from

  $-\infty < x < +2.35 = 99.1\%$.
- If $x < +1.96$,
- the EXCEL's *=NORMDIST(1.96,0,1,TRUE)* gives 0.975, i.e. the area under the curve for x>+1.96 = 0.25%.



Shaded area = Probability ***P***

	|z|	*P*	|z|	*P*	|z|	*P*
	0.00	0.0000	1.00	0.3413	2.00	0.4773
	0.10	0.0398	1.10	0.3643	2.10	0.4821
	0.20	0.0793	1.20	0.3849	2.20	0.4861
	0.30	0.1179	1.30	0.4032	2.30	0.4893
	0.40	0.1554	1.40	0.4192	2.40	0.4918
	0.50	0.1915	1.50	0.4332	2.50	0.4938
	0.60	0.2258	1.60	0.4452	2.60	0.4953
	0.70	0.2580	1.70	0.4554	2.70	0.4965
	0.80	0.2881	1.80	0.4641	2.80	0.4974
	0.90	0.3159	1.90	0.4713	3.00	0.4987

	|z|	*P*	|z|	*P*	|z|	*P*	|z|	*P*
	1.90	0.4713	1.94	0.4738	1.94	0.4738	1.98	0.4761
	1.91	0.4719	1.95	0.4744	1.95	0.4744	1.99	0.4767
	1.92	0.4726	1.96	0.4750	1.96	0.4750	2.00	0.4772
	1.93	0.4732	1.97	0.4756	1.97	0.4756		

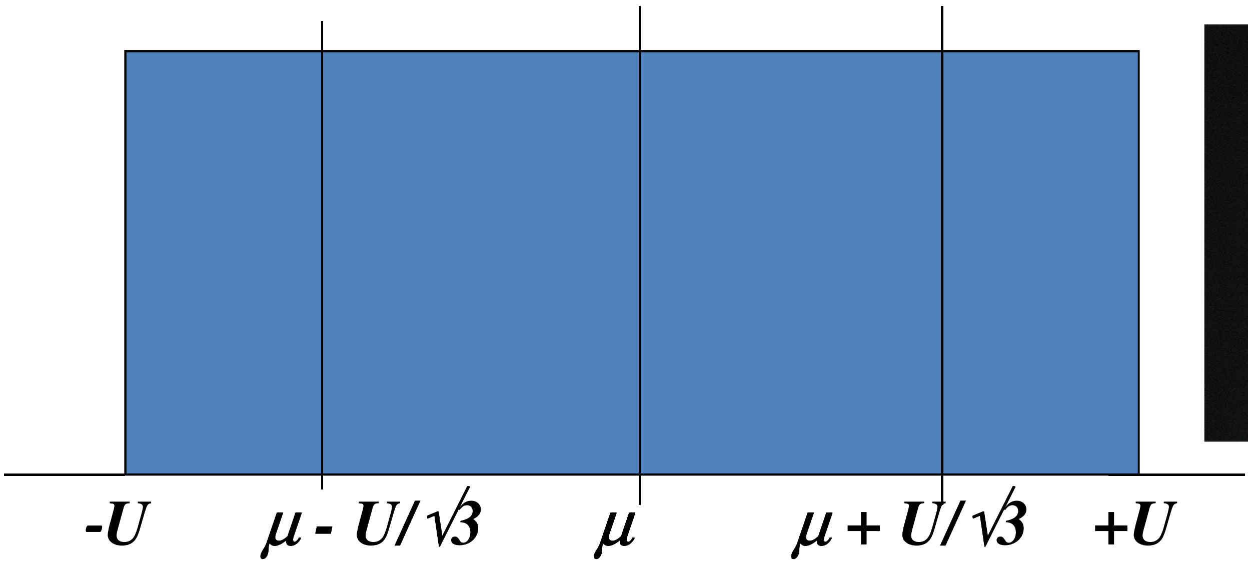
x results （%）	|z|	2-sided area	Probability *P*
$\mu \pm \sigma$	1	0.6826	68.26
$\mu \pm 2\sigma$	2	0.9544	95.44
$\mu \pm 1.96\sigma$	1.96	0.9500	95.00
$\mu \pm 3\sigma$	3	0.9974	99.74

Probability Distribution Functions

- **Normal Distribution**
- **Example**:
- Given a reference weight = 50.0012 gm $\pm$ 0.0009 gm reported by a calibration laboratory with 95% confidence
- The **standard uncertainty** in the form of std deviation = 0.0009/1.96 or 0.0004(6) gm

Probability Distribution Functions

- **Rectangular Probability Distribution**
- It is used when uncertainties are given by ***maximum bound*** within which all values are ***equally probable*** (i.e. P=1).
- Factor = $\sqrt{3}$

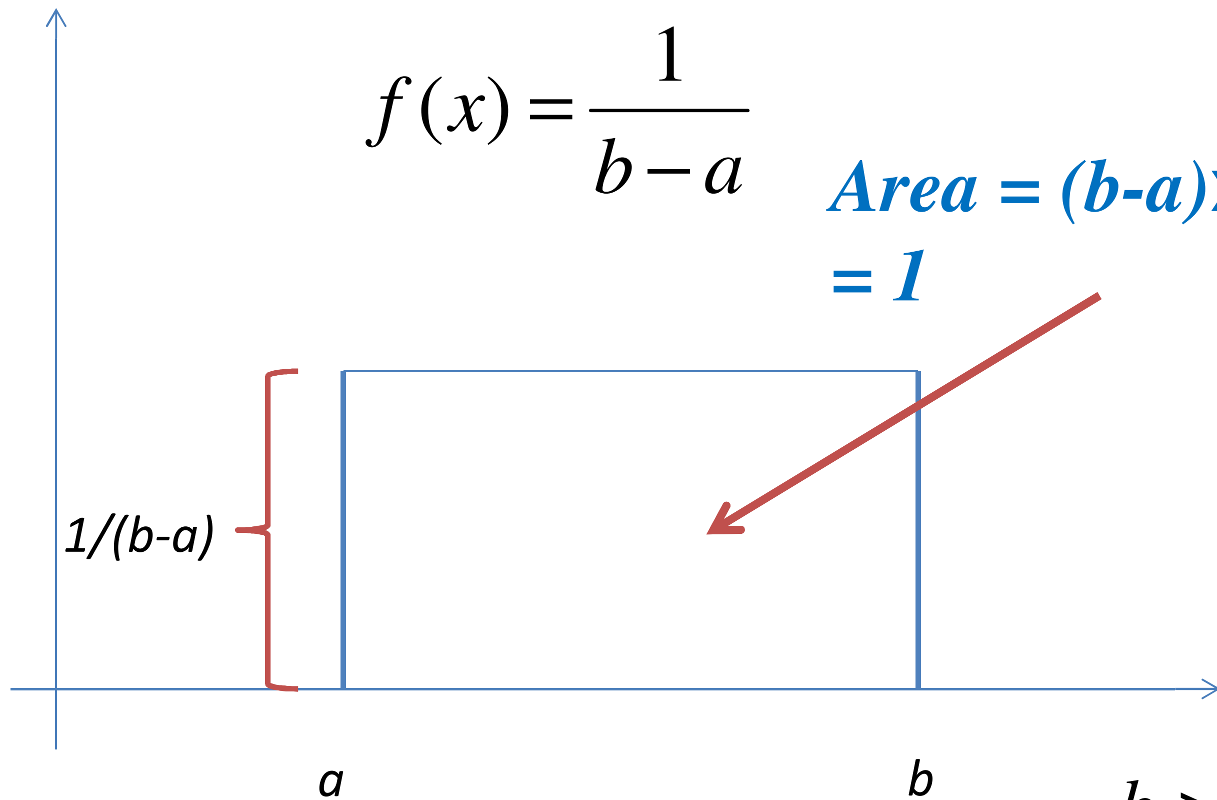


Factor √3 for Rectangular Distribution?

- $y=f(x)$

$$f(x)=\frac{1}{b-a}$$

Area = (b-a)*x*(1/(b-a)) = 1



$$b \geq x \geq a$$

$$y = f(x) = \frac{1}{b-a} \qquad b \geq x \geq a$$

Expected value $\mu = E(X) = \frac{b+a}{2}$ obtained as below:

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{2(b-a)}\left[x^2\right]_a^b$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}$$

By definition of **variance**: $V(X) = E(X^2) - [E(X)]^2$

$$\sigma^2 = V(X) = \int_a^b x^2 . \frac{1}{b-a} dx - \left(\frac{b+a}{2}\right)^2 = \frac{1}{3(b-a)} \left[x^3\right]_a^b - \left(\frac{b+a}{2}\right)^2$$

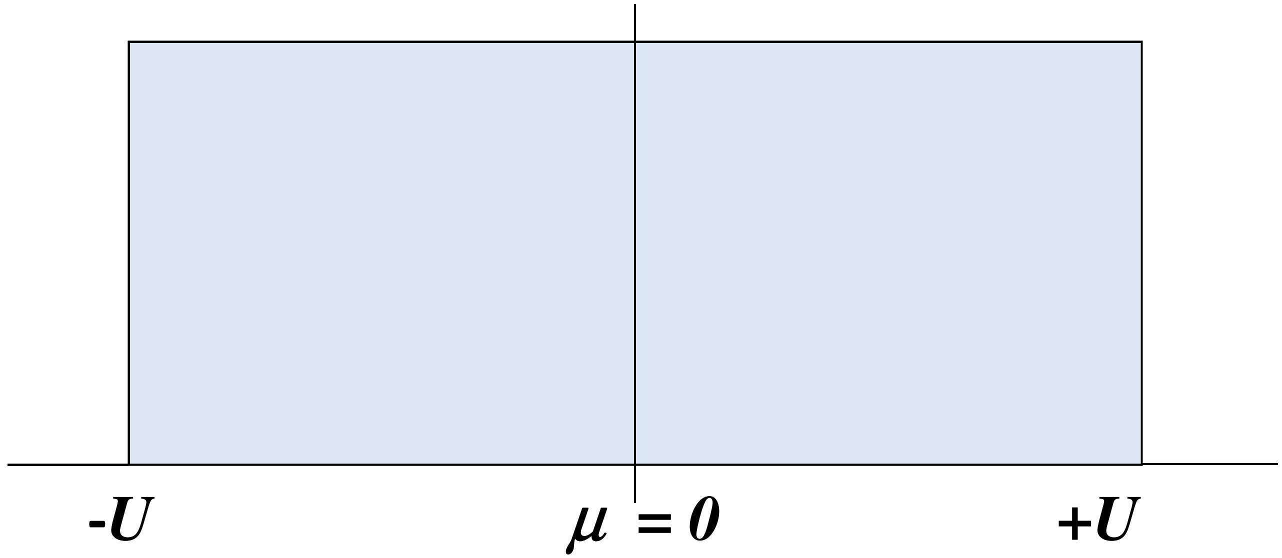
$$= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4}$$

$$= \frac{(b-a)^2}{12}$$

Therefore:
standard deviation

$$\sigma = \frac{b-a}{2\sqrt{3}}$$

- If

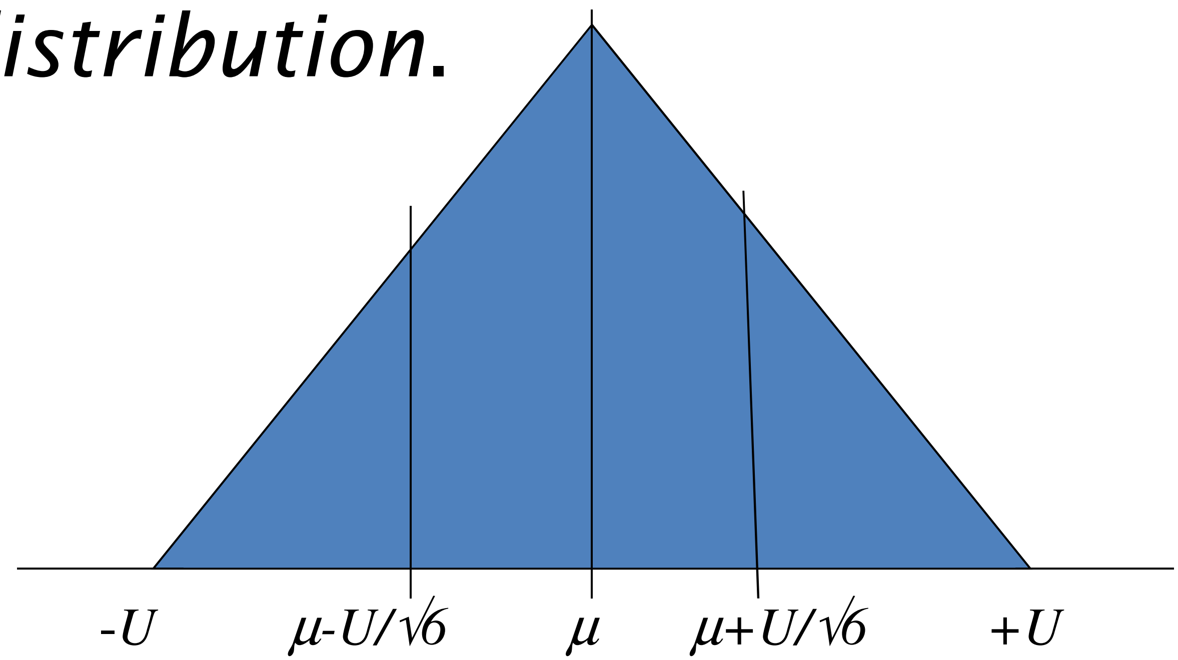


$$f(x) = \frac{1}{U-(-U)} = \frac{1}{2U}$$

$$\sigma = \frac{U-(-U)}{2\sqrt{3}} = \frac{U}{\sqrt{3}}$$

Probability Distribution Functions

- **Triangular Probability Distribution**
- Most of the measured values are much *more likely to occur near the centre point of the distribution*.
- Factor = $\sqrt{6}$



Use of Probability Functions

- **Example**:
- A glassware manufacturer states in his product catalog that his Class B 100 ml volumetric flask has a volume uncertainty of 0.12 ml, without stating its confidence limit.
- Hence, the std uncertainty expressed as std deviation = $0.12/\sqrt{3} = 0.069$ ml. Why?

Another Example…

- Given the purity of an analytical solvent, Methylene chloride as >99%.
- What is the standard uncertainty of the purity of this solvent?
- *Answer:*
- The std uncertainty expressed as standard deviation = $(100-99)/(2 * \sqrt{3})$. **Why?**