

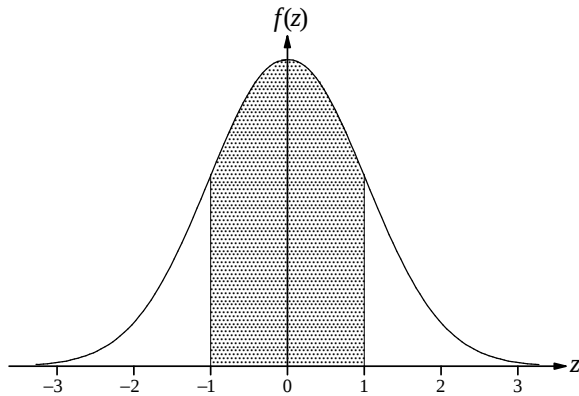
Figure 5.1. PROBABILITY MODELS: Selected Areas Under the Gaussian P.d.f.

The diagrams and equations in this Figure show areas under the standard Gaussian p.d.f. [$f(z)$] that are useful in probability calculations and statistical methods.

- * The upper three diagrams illustrate what is sometimes called the *68–95–99.7 rule*; they involve *integer* numbers of standard deviations from the mean.
- * The three diagrams at the bottom of this side of the Figure refer to ‘round’ values (*viz.* 90%, 95% and 99%) for the *central* area (as a percentage of the total area); they are relevant to finding a confidence interval, an idea introduced in Chapter 13.

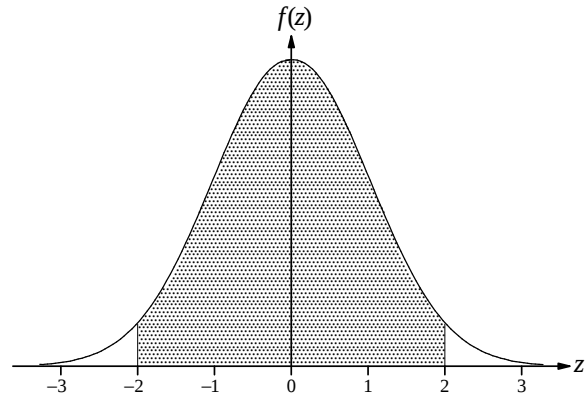
Shaded area: mean ± 1 standard deviation

$$0.6827 = 68.27\% \approx \mathbf{68\%}$$



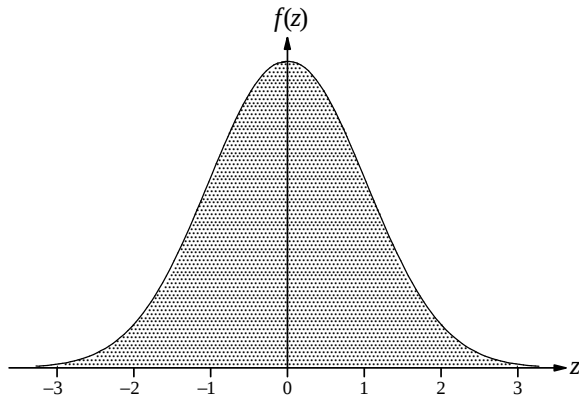
Shaded area: mean ± 2 standard deviations

$$0.9545 = 95.45\% \approx \mathbf{95\%}$$



Shaded area: mean ± 3 standard deviations

$$0.9973 = 99.73\% \approx \mathbf{99.7\%}$$



Other central (C) and tail (2T) areas:

- **mean ± 3 standard deviations:**

$$\text{C: } 0.997\ 300$$

$$\text{2T: } 0.002\ 700 \approx 3 \text{ in } 10^3$$

- **mean ± 4 standard deviations:**

$$\text{C: } 0.999\ 936\ 66$$

$$\text{2T: } 0.000\ 063\ 34 \approx 6 \text{ in } 10^5$$

- **mean ± 5 standard deviations:**

$$\text{C: } 0.999\ 999\ 4268$$

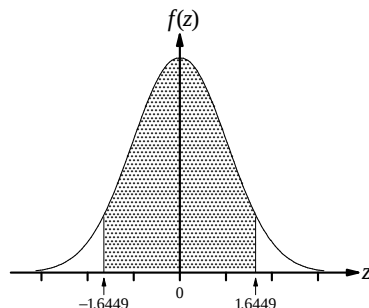
$$\text{2T: } 0.000\ 000\ 5732 \approx 6 \text{ in } 10^7$$

- **mean ± 6 standard deviations:**

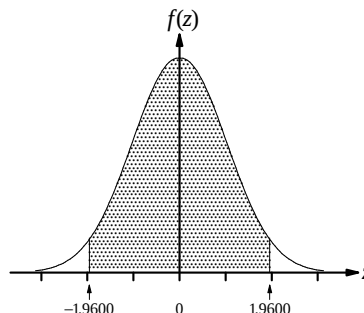
$$\text{C: } 0.999\ 999\ 998\ 0268$$

$$\text{2T: } 0.000\ 000\ 001\ 9732 \approx 2 \text{ in } 10^9$$

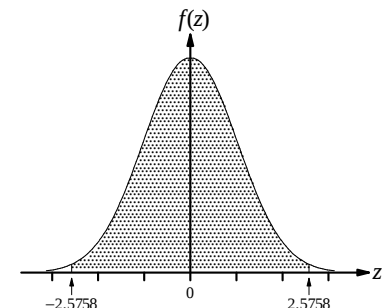
Shaded area: central 90%
mean ± 1.6449 standard deviations



Shaded area: central 95%
mean ± 1.9600 standard deviations



Shaded area: central 99%
mean ± 2.5758 standard deviations



The probability expressions equivalent to the *central* areas in the diagrams and equations given overleaf on page 5.1 are as follows, where the random variable $Z \sim G(0, 1)$ and the random variable $Y \sim G(\mu, \sigma)$:

$\Pr(-1 < Z \leq 1) = 0.6827$	$\Pr(\mu - \sigma < Y \leq \mu + \sigma) = 0.6827$
$\Pr(-2 < Z \leq 2) = 0.9545$	$\Pr(\mu - 2\sigma < Y \leq \mu + 2\sigma) = 0.9545$
$\Pr(-3 < Z \leq 3) = 0.9973$	$\Pr(\mu - 3\sigma < Y \leq \mu + 3\sigma) = 0.9973$
$\Pr(-4 < Z \leq 4) = 0.999\,936\,66$	$\Pr(\mu - 4\sigma < Y \leq \mu + 4\sigma) = 0.999\,936\,66$
$\Pr(-5 < Z \leq 5) = 0.999\,999\,4268$	$\Pr(\mu - 5\sigma < Y \leq \mu + 5\sigma) = 0.999\,999\,4268$
$\Pr(-6 < Z \leq 6) = 0.999\,999\,998\,0268$	$\Pr(\mu - 6\sigma < Y \leq \mu + 6\sigma) = 0.999\,999\,998\,0268$
$\Pr(-1.6449 < Z \leq 1.6449) = 0.90$	$\Pr(\mu - 1.6449\sigma < Y \leq \mu + 1.6449\sigma) = 0.90$
$\Pr(-1.9600 < Z \leq 1.9600) = 0.95$	$\Pr(\mu - 1.9600\sigma < Y \leq \mu + 1.9600\sigma) = 0.95$
$\Pr(-2.5758 < Z \leq 2.5758) = 0.99;$	$\Pr(\mu - 2.5758\sigma < Y \leq \mu + 2.5758\sigma) = 0.99.$

The corresponding (two)-tail probability expressions are:

$\Pr(Z > 1) = 0.3173$	$\Pr(Y > \mu + \sigma) = 0.3173$
$\Pr(Z > 2) = 0.0455$	$\Pr(Y > \mu + 2\sigma) = 0.0455$
$\Pr(Z > 3) = 0.0027$	$\Pr(Y > \mu + 3\sigma) = 0.0027$
$\Pr(Z > 4) = 0.000\,063\,34$	$\Pr(Y > \mu + 4\sigma) = 0.000\,063\,34$
$\Pr(Z > 5) = 0.000\,000\,5732$	$\Pr(Y > \mu + 5\sigma) = 0.000\,000\,5732$
$\Pr(Z > 6) = 0.000\,000\,0019732$	$\Pr(Y > \mu + 6\sigma) = 0.000\,000\,0019732$
$\Pr(Z > 1.6449) = 0.10$	$\Pr(Y > \mu + 1.6449\sigma) = 0.10$
$\Pr(Z > 1.9600) = 0.05$	$\Pr(Y > \mu + 1.9600\sigma) = 0.05$
$\Pr(Z > 2.5758) = 0.01;$	$\Pr(Y > \mu + 2.5758\sigma) = 0.01.$

NOTES: 1. Unless context dictates otherwise, we write $\Pr(a < Z \leq b)$, the probability the random variable Z (say) takes values between a and b , so as to exclude the lower end-point of the interval $(a, b]$ and include the (finite) upper end-point.

- This convention maintains consistency when we use our definition of the cumulative distribution function $F(y) = \Pr(Y \leq y)$ to find a probability (e.g., see Table 5.3.1 on page 5.8 in Figure 5.3).

2. In the *Six Sigma* process improvement program, it is said ‘ 6σ has been achieved’ when the defect rate does not exceed 3.4 defects per million opportunities. Curiously, this value is the tail area beyond 4.5 standard deviations from the mean; i.e., $\Pr(Z > 4.5) = \Pr(Y > \mu + 4.5\sigma) = 0.000\,003\,398$.

- The 4.5 standard deviations is the difference between 6σ (the idea from which the program takes its name) and an assumed 1.5σ shift of the process centre from the mean.

- ① The vertical axis in each of the six diagrams shown overleaf is labelled $f(z)$; give both the *name* and the *equation* for $f(z)$.
- ② How would the *values* of the central probabilities given above be affected if the $<$ signs in the inequalities were replaced by $\leq$? Explain briefly.
 - Answer the same question if the $\leq$ signs were replaced by $<$ signs.
- ③ How would the *values* of the tail probabilities given above be affected if the $>$ signs in the inequalities were replaced by $\geq$? Explain briefly.
- ④ Why are the absolute value signs used in the two-tail probability expressions above? Rewrite one of these expressions without using absolute value signs.
- ⑤ Write the *one-tail* probability expressions corresponding to the two-tail expressions given above.
- ⑥ Sketch the diagrams and write the probability expressions corresponding to the central 80% and 98% of the area under the Gaussian p.d.f.