Figure 5.3. NORMAL DISTRIBUTIONS: Selected Areas Under the P.d.f.

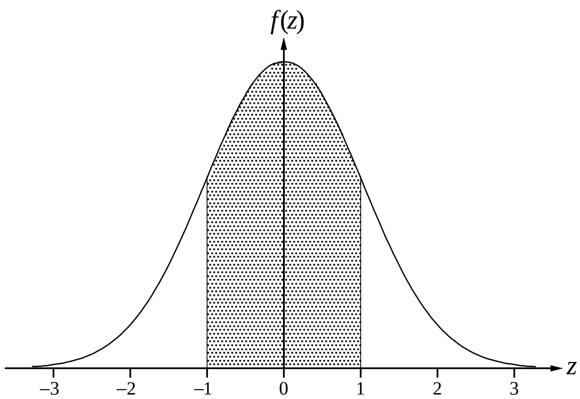
The diagrams and equations in this Figure show areas under the standard normal p.d.f. [$f(z)$] that are useful in probability calculations and statistical methods.

∗ 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 Part 6.

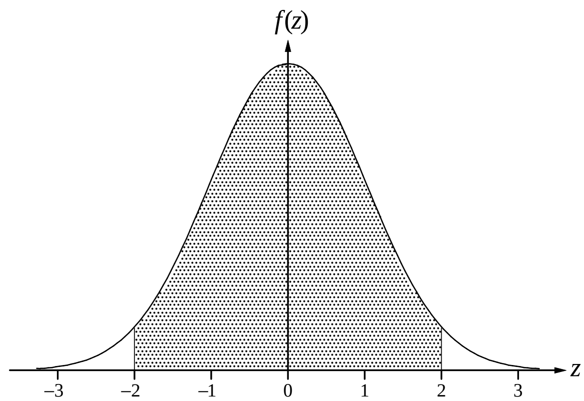
Shaded area: mean ± 1 standard deviation

0.6827 = 68.27% $\simeq$ **68%**



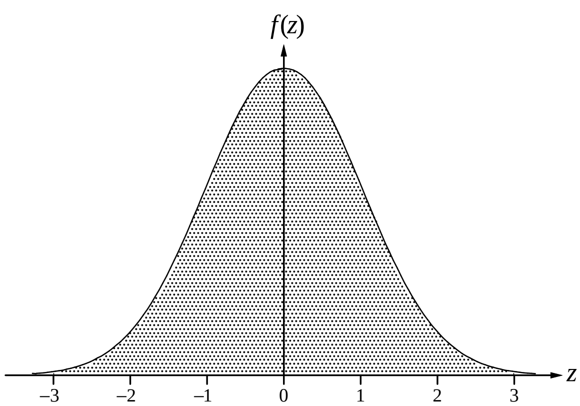
Shaded area: mean ± 2 standard deviations

0.9545 = 95.45% $\simeq$ **95%**



Shaded area: mean ± 3 standard deviations

0.9973 = 99.73% $\simeq$ **99.7%**

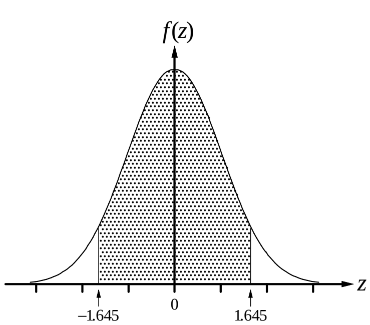


Other central (C) and tail (2T) areas:

- **mean ± 3 standard deviations**:
  - **C**: 0.997 300
  - **2T**: 0.002 700 $\simeq$ 3 in 10^3
- **mean ± 4 standard deviations**:
  - **C**: 0.999 936 66
  - **2T**: 0.000 063 34 $\simeq$ 6 in 10^5
- **mean ± 5 standard deviations**:
  - **C**: 0.999 999 4268
  - **2T**: 0.000 000 5732 $\simeq$ 6 in 10^7
- **mean ± 6 standard deviations**:
  - **C**: 0.999 999 998 0268
  - **2T**: 0.000 000 001 9732 $\simeq$ 2 in 10^9

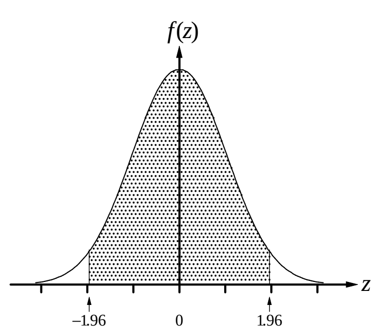
Shaded area: central 90%

mean ± 1.6449 standard deviations



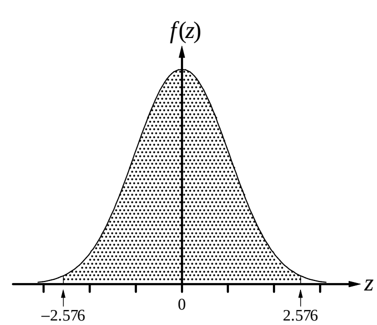
Shaded area: central 95%

mean ± 1.9600 standard deviations



Shaded area: central 99%

mean ± 2.5758 standard deviations



(*continued overleaf*)

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

$\Pr(-1 < Z \le 1) = 0.6827$	$\Pr(\mu - \sigma < Y \le \mu + \sigma) = 0.6827$
$\Pr(-2 < Z \le 2) = 0.9545$	$\Pr(\mu - 2\sigma < Y \le \mu + 2\sigma) = 0.9545$
$\Pr(-3 < Z \le 3) = 0.9973$	$\Pr(\mu - 3\sigma < Y \le \mu + 3\sigma) = 0.9973$
$\Pr(-4 < Z \le 4) = 0.999\,936\,66$	$\Pr(\mu - 4\sigma < Y \le \mu + 4\sigma) = 0.999\,936\,66$
$\Pr(-5 < Z \le 5) = 0.999\,999\,4268$	$\Pr(\mu - 5\sigma < Y \le \mu + 5\sigma) = 0.999\,999\,4268$
$\Pr(-6 < Z \le 6) = 0.999\,999\,998\,0268$	$\Pr(\mu - 6\sigma < Y \le \mu + 6\sigma) = 0.999\,999\,998\,0268$
$\Pr(-1.645 < Z \le 1.645) = 0.90$	$\Pr(\mu - 1.645\sigma < Y \le \mu + 1.645\sigma) = 0.90$
$\Pr(-1.96 < Z \le 1.96) = 0.95$	$\Pr(\mu - 1.96\sigma < Y \le \mu + 1.96\sigma) = 0.95$
$\Pr(-2.576 < Z \le 2.576) = 0.99$;	$\Pr(\mu - 2.576\sigma < Y \le \mu + 2.576\sigma) = 0.99$.

The corresponding (*two*)-*tail* probability expressions are:

$\Pr(\lvert Z\rvert > 1) = 0.3173$	$\Pr(\lvert Y\rvert > \mu + \sigma) = 0.3173$
$\Pr(\lvert Z\rvert > 2) = 0.0455$	$\Pr(\lvert Y\rvert > \mu + 2\sigma) = 0.0455$
$\Pr(\lvert Z\rvert > 3) = 0.0027$	$\Pr(\lvert Y\rvert > \mu + 3\sigma) = 0.0027$
$\Pr(\lvert Z\rvert > 4) = 0.000\,063\,34$	$\Pr(\lvert Y\rvert > \mu + 4\sigma) = 0.000\,063\,34$
$\Pr(\lvert Z\rvert > 5) = 0.000\,000\,5732$	$\Pr(\lvert Y\rvert > \mu + 5\sigma) = 0.000\,000\,5732$
$\Pr(\lvert Z\rvert > 6) = 0.000\,000\,001\,9732$	$\Pr(\lvert Y\rvert > \mu + 6\sigma) = 0.000\,000\,001\,9732$
$\Pr(\lvert Z\rvert > 1.645) = 0.10$	$\Pr(\lvert Y\rvert > \mu + 1.645\sigma) = 0.10$
$\Pr(\lvert Z\rvert > 1.96) = 0.05$	$\Pr(\lvert Y\rvert > \mu + 1.96\sigma) = 0.05$
$\Pr(\lvert Z\rvert > 2.576) = 0.01$;	$\Pr(\lvert Y\rvert > \mu + 2.576\sigma) = 0.01$.

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

- This convention maintains consistency when we use our definition of the cumulative distribution function $F(z)$ in Figure 5.9 to find a probability.

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 ≤? Explain briefly.

- Answer the same question if the ≤ 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 ≥? Explain briefly.

4 Why are the absolute value signs used in the two-tail probability expressions above? Rewrite one of these expressions with*out* 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 normal p.d.f.