For Normal (Gaussian) Distributions, given two independent data sets,

$$x \pm \delta x, y \pm \delta y ,$$

then combining the two data sets we get

$$x \pm y \pm \sqrt{\delta x^2 + \delta y^2} ,$$

$$\{xy, x/y\} \pm \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2}$$

.

The error when adding (subtracting) the two sets can be understood by considering a Gaussian in two independent variables–a 2 dimensional Guassian–centered at the origin. Since the error in x and the error in y are the deviation from zero in orthogonal directions, they combine through the Pythagorean relation.

The Normal Distribution Function,

$$p_u(-\infty, x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

is the probability that a point taken from a Normal Distribution, with variance σ and mean μ, is between $-\infty$ and x.

(Note that here we have added the subscript u for "unscaled", while we will reserve the notation $p(x)$ for the "scaled", Standard Normal Distribution Function.) The Normal Distribution Function is rescaled to become the Standard Normal Distribution Function through the change of variables,

$$\tilde{t} = \frac{t-\mu}{\sigma} \to \tilde{x} = \frac{x-\mu}{\sigma}, \tilde{d}t = \frac{dt}{\sigma} .$$

Dropping the tildes we get the Standard Normal Distribution Function with mean zero and x rescaled so it is now the number of standard deviations from the mean:

$$p(-\infty, x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt .$$

Also note that since the integrand is normalized and an even function, $p(-\infty, \infty) = 1$ and $p(-\infty, 0) = \frac{1}{2}$. We have defined the standard normal distribution function to allow us to vary both limits of integration. Doing this gives

$$p(-x, x) = 2p(0, x) = \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) ,$$

where $\mathrm{erf}(x)$ is the error function,

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

Note that $2p(0, x)$ is the probability of finding a point within x standard deviations from the mean while $1 - 2p(0, x)$ is the probability of finding a point outside x standard deviations from the mean. The values for both $p(0, x)$ and $\mathrm{erf}(x)$ can be found in tables (such as the standard Abramowitz and Stegun, Handbook of Mathematical Functions) and are often included as functions in math programs, such as Mathematica. (In Mathematica use Erf[x], and don't forget about the square root of 2.)

For this problem, find the difference between the means of the two data sets and divide by their combined error; call this a. The probability that the two data sets are NOT from the same distribution is $2p(0, a)$. Equivalently, the probability that they are from the same distribution is $1 - 2p(0, a)$. Note the interpretations of the limits $a \to 0$ and $a \to \infty$.

Reference: `http://web.engr.illinois.edu/~schleife/teaching/mse485/dokuwiki/doku.php?id=probability_help`