Two Samples: $X_1, \dots, X_q, Y_1, \dots, Y_r$ such that $X \sim N(\mu_x, \sigma^2), Y \sim N(\mu_Y, \sigma^2), where\ \mu_x - \mu_y = \delta > 0$

First, note the fact that finding $P(X_{(q)} > Y_{(r)})$ is equivalent to finding the probability that the largest value from the combined sample originally came from the X distribution, since the max must be $X_{(q)}$ or $Y_{(r)}$ (the respective maximums from the X, Y samples).

Notice, now, that the only difference between our distributions is the means. Because of this, we can express f_Y as f_x with a location shift ($f_y(y) \to f_x(y + \delta)$). More formally,

Result 1

$$f_y(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\left(y - \mu_y\right)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(y - (\mu_x - \delta))^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{((y + \delta) - \mu_x)^2}{2\sigma^2}\right)$$

$$\to^d f_x(y + \delta)$$

It then follows naturally that $F_y(y) = F_x(y + \delta)$

To find the distribution of the maximum value of either distribution (we will start with X first), we need the CDF, which we will denote $F_x(x)$. Notice that, if $X_{(q)}$ is the maximum value of a sample of size q, then for all other X, $X_i \leq X_{(q)}$. We would of course express this in terms of probability as $P\left(X \leq X_{(q)}\right)^{q-1}$ or $F_x\left(X_{(q)}\right)^{q-1}$, since we assume the samples are collected independently. In addition to this, we also need the probability of observing this value, and then multiply it by the sample size (since, theoretically, any one of the q samples could be the max). So the pdf of the maximum X value is:

$$f_V(v) = q f_x(v) * \left(F_x(v)\right)^{q-1}, -\infty < v < \infty\ where\ v = X_{(q)}$$

Similarly, for Y:

$$f_W(w) = r f_y(w) * \left(F_y(w)\right)^{r-1}, -\infty < w < \infty\ w = Y_{(r)}$$

Using Result 1, we can represent $f_w(w)$ as

$$f_w(w) = r f_x(w + \delta) F_x(w + \delta)^{r-1}$$

We now have distribution functions for the maximum value of each sample. Next, lets get the joint distribution of v and w. Since we are assuming independence, this is simply the product of the two pdfs:

$$f_{v,w}(v, w) = f_v(v) * f_w(w) = qr f_x(v) f_x(w + \delta) F_x(v)^{q-1} F_x(w + \delta)^{r-1}$$

We can now attempt to solve for the probability that $v > w$ via integration:

$$\int_{-\infty}^{\infty}\int_{w}^{\infty} qrf_x(v)f_x(w+\delta)F_x(v)^{q-1}F_x(w+\delta)^{r-1}\ dv\ dw$$

$$= \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}\int_{w}^{\infty} qf_x(v)F_x(v)^{q-1}\ dv\ dw$$

$$= \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}\left(\int_{w}^{\infty} q\,(u)^{q-1}du\right)dw, \qquad where\ u = F_x(v), du = f_x(v)dv$$

$$= \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}\,(F_x(v)^q|_w^{\infty}dw = \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}\,(1-F_x(w)^q)dw$$

$$= 1 - \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}F_x(w)^q\ dw = 1 - \int_{-\infty}^{\infty} rf_x(w+\delta)F_x(w+\delta)^{r-1}F_x(w)^q\ dw$$

Here is where we begin to run into trouble. Under certain conditions, the above is solvable analytically, expressed in terms of a normal cdf. However, unless q,r=1 (meaning we are only comparing between samples of 1), this **cannot be solved analytically,** at least to the point that it cannot be expressed in familiar functions. To proceed, we will have to look at each individual case separately. To do this, I will solve this integral numerically for each case and perform a monte carlo simulation based on the original problem to confirm the answer.

The best way to simulate an integral with infinite bounds is with a method called importance sampling. We start with the integral and convert it to an expectation problem – something we can do easily via simulation. So have the following:

$$I = \int_{-\infty}^{\infty} g(x)\,dx$$

Where $g(x)$ is the function defined above ($g(w) = rf_x(w+\delta)F_x(w+\delta)^{r-1}F_x(w)^q$). Now, we transform this using a function on the same bounds that can be simulated from easily, which mimics the form of $g(x)$. For our purposes, we will use $f \sim N(1010,10)$, which can be easily simulated in R. So our integral becomes:

$$I = \int_{-\infty}^{\infty} \frac{g(x)}{f(x)} f(x)dx = E\left(\frac{g(x)}{f(x)}\right)$$

To simulate this at 100,000 repetitions, we generate 100,000 random numbers from our $f(x)$ distribution. From there, we calculate $g(x)/f(x)$ for those 100,000 numbers and take the average. This is our approximation of the integral. From there, we subtract it from 1 to get the probability that the largest value for our distributions came from X, or the distribution with the larger mean.

See the below table for the results from the numerical integration and the simulation for the specified inputs:

q	MC_Integration	MC_Simulation
1	0.1899605	0.18684
2	0.3670318	0.36786
3	0.5370004	0.53951
4	0.6989813	0.69772
5	0.8534716	0.85225

I have put the R code in the appendix. It should be fairly straightforward to update the code for different specs- but depending on how much the means differ by you may need to change the sampling distribution. Note that the results may be slightly different since I did not use a random seed.

CODE APPENDIX

```
##Integration Code


g1 <- function(w,del,q,r,mu,s){r*dnorm(w+del,mu,s)*(pnorm(w+del,mu,s)^(r-1))*(pnorm(w,mu,s)^q)}
g1 <- Vectorize(g1)


##Simulation
Results <- data.frame(q=seq(5),MC_Integration=0,MC_Simulation=0)
for (i in 1:5){
        M <- 100000
        ##Integration
        q <- i
        r <- 6-i
        x <- rnorm(M,1010,10)
        y <- g1(x,1,q,r,1000,10)
        Results[i,2] <- 1-mean(y/dnorm(x,1010,10))
        ##Simulation
        count <- 0
        for (j in 1:M){
                X <- rnorm(q,1001,10)
                Y <- rnorm(r,1000,10)
                if (max(X) > max(Y)){ count = count + 1}
                }
        Results[i,3] <- count/M
}
Results
```