6-2009

Confidence Intervals for the Ratio of Two Exponential Means with Applications to Quality Control

James Albert Polcer,III
Western Kentucky University, james.polcer@wku.edu

Confidence Intervals for the Ratio of Two Exponential Means With Applications to Quality Control *

James Polcer

June 23, 2009

Project Director:	Dr. Jonathan Quiton
Committe Members:	Dr. Di Wu
	Dr. Melanie Autin

Abstract

We considered the problem of statistical quality control based on the ratio of two population means. We restrict the discussion for two exponential rates, which are commonly used for modeling failure times of components, machines, or systems. Closed form expressions via the moment generation function (MGF) technique will be presented, and numerical examples will be shown using engineering data sets.

1 Introduction

Statistical control charts monitor a single process over time. We would like to take an extra step and monitor two processes over time. This will allow us to use the ratio of the two means and analyze the data we see. Exponential distributions measure failure times. Are the processes failing at a constant rate? Is one process failing at a greater

*James Polcer is a B.S. Mathematics student, Western Kentucky University. Dr. Jonathan Quiton, Dr. Di Wu, and Dr. Melanie Autin are faculty of Mathematics Department, Western Kentucky University. Mr. Polcer's research is suppported by the Kentucky EPSCoR Research Startup Fund RSF-031-06.

rate than the other? We chose to use the exponential distribution because exponentially distributed populations are common in engineering applications.

For illustration purposes, we generated a sample of 6 observations from Exponential ($\theta_2 = 200$) and 8 observations from Exponential ($\theta_1 = 100$). Let the first set be group "y" and the second group as "x":

Table 1: Generated data for illustration purposes

sample size	data
$n_2 = 6$	$y = \{30, 946, 53, 188, 96, 20\}$
$n_1 = 8$	$x = \{9, 183, 30, 193, 37, 33, 229, 38\}$

Suppose the only information we know is what appears on the table. We are interested in constructing a $(1 - \alpha) \times 100$ confidence interval for $\rho = \frac{\theta_2}{\theta_1}$. Furthermore, we will extend this method to quality control setting where we are interested in spotting "out-of-control" batches based on the confidence intervals of the ratios.

2 Method

For brevity of notation, the subscript i for the random variables are omitted since we are focusing on one batch. Let $\mathbf{y} = \{y_1...y_n\} \sim Exp(\theta_2)$ and $\mathbf{x} = \{x_1...x_n\} \sim Exp(\theta_1)$ where $\mathbf{y}$ and $\mathbf{x}$ are the two random vectors generated from exponential rates θ_2, θ_1 respectively.

When making a confidence interval the first, and most crucial, part is finding a pivotal quantity (or pivot for short). By definition, a pivotal quantity is a function of the data and the parameter(s) such that the distribution of that pivotal quantity does not depend on the parameter(s). In this setting, our pivot is of the form:

$$Z = \frac{\frac{\overline{x}}{\theta_1}}{\frac{\overline{y}}{\theta_2}}.$$

Observe that Z is a function of the parameters and the data, satisfying the first criterion. For the second criterion, we need to show that the density of Z does not depend on the parameters.

2.1 What is the density of our pivot?

To derive our pivot, we first state and prove the following lemma:

Lemma 1 *Let* $\mathbf{x} = \{x ... x_n\}$ *be a collection of independent and identically distributed random variables with* $x_i \sim Exp(\theta_1)$ *for* $i = 1, 2, \cdots, n$. *Then* $\frac{\overline{x}}{\theta_1} \sim Gamma(n_1, \frac{1}{n_1})$.

The method of proof uses the Moment Generating Function (MGF) technique. This technique is based on a theorem showing that a random variable X can be uniquely identified by its Moment Generating Function, $M_X(t) = E\left(e^{tX}\right)$. This basically means that if I can find the MGF of some form, then I know the distribution it comes from. So, supposing that $x_i \sim$ Exponential(θ_1) with the density

$$f(x_i) = \frac{1}{\theta_1} \exp\left(\frac{-x_i}{\theta_1}\right),$$

standard statistical theory (c.f. Hoog and Tanis,2006) shows that the MGF of an exponential distribution is:

$$M_{X_i}(t) = \frac{1}{1 - \theta_1 t}$$

Using MGF properties, I can find the distribution of $\sum_{i=1}^{n}(x_i)$ via the moment generating function:

$$\begin{aligned}
M_{\sum_{i=1}^{n} X_i}(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\
&= \prod_{i=1}^{n}\left[\frac{1}{1-\theta_1(t)}\right] \\
&= \frac{1}{1-\theta_1(t)} \times \frac{1}{1-\theta_1(t)} \times \frac{1}{1-\theta_1(t)} \cdots \\
&= \frac{1}{(1-\theta_1 t)^n} \qquad (1)
\end{aligned}$$

From the above result, $M_{\sum_{i=1}^{n}(X_i)}(t)$ turns out to be an MGF of a Gamma random variable with parameters $(\alpha = n, \beta = \theta_1)$. Consequently, the MGF of the mean of exponentials turns out to be:

$$\begin{aligned}
M_{\bar{X}}(t) &= M_{\frac{\sum_{i=1}^{n} X_i}{n}}(t) \\
&= M_{\sum_{i=1}^{n} X_i}\left(\frac{t}{n}\right) \\
&= \frac{1}{\left(1-\frac{\theta_1}{n}t\right)^n} \qquad (2)
\end{aligned}$$

which is another Gamma random variable with parameters $\left(\alpha = n, \beta = \dfrac{\theta}{n}\right)$. Finally,

$$\begin{aligned}
M_{\frac{\bar{X}}{\theta_1}}(t) &= M_{\bar{X}}\left(\frac{t}{\theta_1}\right) \\
&= \frac{1}{(1-\theta(\frac{t}{n\theta}))^n} \\
&= \frac{1}{(1-\frac{1}{n}t)^n} \qquad (3)
\end{aligned}$$

which is the MGF of a Gamma random variable with parameters $\left(\alpha = n, \beta = \frac{1}{n_1}\right)$. From Lemma 1, we can also show that the ratio $\frac{\bar{Y}}{\theta_2}$ is also a Gamma random variable with parameters $\left(\alpha = n_2, \beta = \frac{1}{n_2}\right)$.

The next step is to find the distribution of the ratio of two gamma random variables. In this case, we will use the transformation technique which is useful for nonlinear transformations such as ratios. Let

$$
\begin{aligned}
W &= \frac{\bar{X}}{\theta_1} \sim Gamma\left(n_1, \frac{1}{n_1}\right) \\
V &= \frac{\bar{Y}}{\theta_2} \sim Gamma\left(n_2, \frac{1}{n_2}\right)
\end{aligned}
$$

Let (Z, t) be a new pair of variables such that $Z = W/V$ and $t = V$. Transformation technique requires that we specify the same number of new variables even if our random variable of interest is only Z. With $V = t$ and $W = Zt$, we form our Jacobian matrix, which is a matrix of partial derivatives of V and W with respect to the new variables Z and t:

$$
J = \begin{bmatrix} \frac{\partial v}{\partial t} = 1 & \frac{\partial v}{\partial z} = 0 \\ \\ \frac{\partial w}{\partial t} = z & \frac{\partial w}{\partial z} = t \end{bmatrix};
$$

and the determinant of this matrix is $|J| = t$. We now have one more step before finding the density of our pivot. The rule behind the transformation technique is that since we are using two or more random variables, we will wind up with a joint density. This is perfectly fine because we can then marginalize the unwanted variables and just focus on Z. Once we do that we will have our desired distribution of the pivot. Here is our joint density from the transformation technique:

$$\begin{aligned}
g(Z,t) &= f_w(Zt) * f_v(t)|J| \\
&= \left[\frac{1}{\Gamma(n_1)(\frac{1}{n_1})^{n_1}}(Zt)^{n_1-1}\exp\left(-\frac{Zt}{\frac{1}{n_1}}\right)\right] \\
&\quad \times \left[\frac{1}{\Gamma(n_2)(\frac{1}{n_2})^{n_2}}(t)^{n_2-1}\exp\left(-\frac{t}{\frac{1}{n_2}}\right)\right]|t| \\
&= \left[\frac{n_1^{n_1}n_2^{n_2}}{\Gamma(n_1)\Gamma(n_2)}z^{n_1-1}\right]t^{n_1+n_2-1}\exp(-(n_1z+n_2)t) \qquad (4)
\end{aligned}$$

Since our interest is on the distribution of Z, marginalize our joint density by integrating out the joint density with respect to all other unwanted variables (in this case, t). This will then give us the density of our pivot:

$$\begin{aligned}
g(z) &= \int_{\forall t} g(z,t)dt \\
&= \left[\frac{n_1^{n_1}n_2^{n_2}}{\Gamma(n_1)\Gamma(n_2)}z^{n_1-1}\right]\left[\frac{\Gamma(n_1+n_2)}{(n_1z+n_2)^{n_1+n_2}}\right] \\
&\quad \times \int_0^\infty \frac{(n_1z+n_2)^{n_1+n_2}}{\Gamma(n_1+n_2)}t^{n_1+n_2-1} \\
&\quad \times \exp(-(n_1Z+n_2)t))dt \\
&= \left[\frac{n_1^{n_1}n_2^{n_2}}{\Gamma(n_1)\Gamma(n_2)}z^{n_1-1}\right]\left[\frac{\Gamma(n_1+n_2)}{(n_1z+n_2)^{n_1+n_2}}\right]. \qquad (5)
\end{aligned}$$

The integration term goes to 1 because we are integrating a gamma density over its domain. For deriving expectations, we can express $g(z)$ as

$$\begin{aligned}
g(z) &= \left(\frac{1}{\frac{\Gamma(n_1)\Gamma(n_2)}{\Gamma(n_1+n_2)}}\right)\left(\frac{n_2}{n_1z+n_2}\right)^{n_2-1} \\
&\quad \times \left(\frac{n_2n_1(n_1z)^{n_1-1}}{(n_1Z+n_2)^{n_1-1}(n_1Z+n_2)^2}\right) \\
&= \frac{n_2n_1}{(n_1z+n_2)^2}\left[\left(\frac{1}{\beta(n_1,n_2)}\right)\left(\frac{n_2}{n_1z+n_2}\right)^{n_2-1}\left(1-\frac{n_2}{n_1z+n_2}\right)^{n_1-1}\right] \qquad (6)
\end{aligned}$$

which has a similar structure to a Beta density. In fact, we can use the Beta density if we change variables. Since g(z) does not depend on the parameters (θ_1, θ_2), then Z is indeed a pivotal quantity.

To verify numerically if our mathematical derivations is correct, we simulated V and W taking 100,000 values from our random variable W and 100,000 values from our random variable V. We then took the ratio of the two and plotted our data, which is indicated by the red bars. We see that empirical distribution fits the theoretical curve nicely.

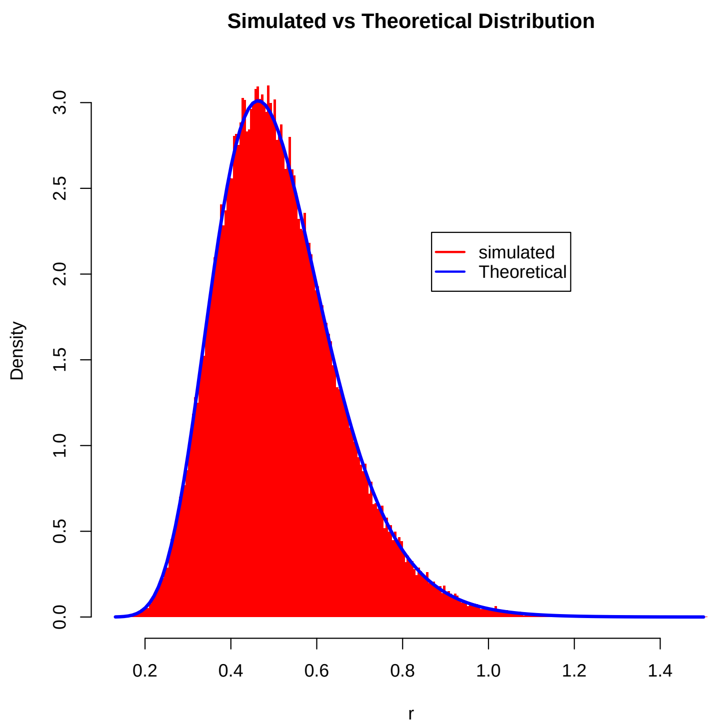

Figure 1: Simulated vs theoretical distribution of the biased ratio estimate: Example 1

2.2 Setting Up Our Confidence Interval

Now we can start working on the confidence interval for the desired ratio: $\frac{\theta_2}{\theta_1}$. Let Z be our pivot with g(z) as its density as specified in (6). We start by setting up the probability statement:

$$P\left(Z_{1-\frac{\alpha}{2}} < \frac{\frac{\overline{x}}{\theta_1}}{\frac{\overline{y}}{\theta_2}} < Z_{\frac{\alpha}{2}}\right) = 1-\alpha \tag{7}$$

Where $Z_{1-\alpha/2}$ and $Z_{\alpha/2}$ are the lower and upper values of Z such the middle probability is $1-\alpha$. The next series of steps is to manipulate the inequality statement such that the middle inequality will be left only with $\frac{\theta_2}{\theta_1}$:

$$P\left((Z_{1-\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}} < \frac{\frac{1}{\theta_1}}{\frac{1}{\theta_2}} < (z_{\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}}\right) = 1-\alpha$$

$$P\left((Z_{1-\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}} < \frac{\theta_2}{\theta_1} < (z_{\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}}\right) = 1-\alpha \tag{8}$$

So, $(Z_{1-\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}}$ is our Lower Bound and $(Z_{\frac{\alpha}{2}})\frac{\overline{y}}{\overline{x}}$ is our Upper Bound. After the oral presentation, we realized that if we substitute $r_1 = 2n_1$ and $r_2 = 2n_2$, the density of Z is actually an F-density with degrees of freedom $(r_1 = 2n_1, r_2 = 2n_2)$. Consequently, $Z_{\alpha/2} = F_{\alpha/2(2n_1,2n_2)}$ and $Z_{1-\alpha/2} = F_{1-\alpha/(2n_1,2n_2)}$. Our original method of getting the critical values are shown in the appendix.

3 Application

3.1 Numerical Example using small samples

We now demonstrate by using the data in Table 1. Suppose we would like to find the 95% confidence interval for the ratio $\dfrac{\theta_2}{\theta_1}$. For this level, $\alpha = 0.05$, $\alpha/2 = 0.025$ and since $n_1 = 8$ and $n_2 = 6$ the critical F values are $F_{0.025(16,12)} = 3.1515$ and $F_{0.975(16,12)} = 0.3461$. Sample means turns out to be $\bar{x} = 222.167$ and $\bar{y} = 93.6$, and so the lower and upper bounds for this ratio are:

$$\begin{aligned} L_B &= 0.3461\left(\frac{222.17}{93.63}\right) = 0.8213 \\ U_B &= 3.1515\left(\frac{222.17}{93.63}\right) = 7.4781 \end{aligned}$$

Thus, the 95% Confidence interval is (0.8213,7.4781). Since the interval contains 1, so we are unable to tell statistically which exponential rate is greater than the other.

3.2 Numerical example for large samples

We would also show how the width of the confidence interval change with increasing sample size. Suppose that we generate two samples similar in setting as in Table 1 but this time we generated 80 samples for the first process (X) and 60 samples from the second process (Y). We would like to find the 95% confidence interval for the ratio θ_2/θ_1.

For this data, we got the sample means of $\bar{y} = 178.18$, and $\bar{x} = 89.79$ and critical values $F_{0.025(160,120)} = 1.4052$ and $F_{0.975(160,120)} = 0.7174$. Thus, the lower and upper bounds are:

$$L_B = 0.7174\left(\frac{178.18}{89.79}\right) = 2.7887$$
$$U_B = 1.4052\left(\frac{178.18}{89.79}\right) = 1.4236$$

So based on this data, the population mean for x_1 is greater than x_2 by a factor of at least 1.4 and at most 2.8 with 95 percent confidence. In both settings the true rate was 2 and this example demonstrate that increasing the sample size will also increase precision as expected.

3.3 Proposed methods for quality control applications

Finally, we would like to propose quality control charts based on the ratio of two exponential means. In quality control setting, this could mean monitoring two independent processes whether the relative performance of the two stays the same or has shifted. For instance, we may be looking at failure rates of sampled components manufactured in Plant A and Plant B monitored weekly for k weeks. To establish notation, let $\mathbf{X}_j = \{X_{1j}, X_{2j}, \cdots, X_{n_1 j}\}$ be the random sample from the jth batch ($j = 1, 2, \cdots, k$) of process X. Similarly, let $\mathbf{Y}_j = \{Y_{1j}, Y_{2j}, \cdots, Y_{n_2 j}\}$ be the random sample from the jth batch for process Y.

3.3.1 Normal Approximation

For large samples, we can use normal approximations to establish the lower and upper bounds of the ratio chart. Suppose that the data is statistically under control, then we can pool the batches as if they are just one sample and estimated a pooled unbiased estimator for ρ,say $\bar{r}$. The next step is to establish the lower and upper bounds for each batch as:

$$LB_j = \bar{r} - 3\sqrt{\widehat{V(\widehat{r_j})}}$$
$$UB_j = \bar{r} + 3\sqrt{\widehat{V(\widehat{r_j})}}$$

Where $\widehat{V(\widehat{r_j})}$ is the estimated variance of the unbiased ratio estimator $\widehat{r_j}$ for the jth batch.

3.3.2 Exact Method

In principle, we use $\widehat{g(z)}$ density to establish the exact lower and upper bounds for $\widehat{r_j}$ where $\widehat{g(z)}$ is the estimated density plugging in the unbiased pooled ratio $\bar{r}$ for the ratio θ_2/θ_1. From this density, we can establish the lower and upper bounds for the jth batch:

$$\begin{aligned} LB_j &= \rho_{j(1-\frac{\alpha}{2})} \\ UB_j &= \rho_{j(\frac{\alpha}{2})} \end{aligned}$$

In both methods the batch is considered out of control if $\widehat{r_j} \notin (\ell_j, U_j)$. We used simulated data in order to compare the performance of the two methods under small and large samples. For large samples, we used a fixed batch sample of 50 for the first group and 60 for the second group. Result show that both exact and normal approximations agree on the bounds. This is not true for the small sample setting as shown in Figure 2. In that case, we used 5 and 6 for the first and second groups, respectively.

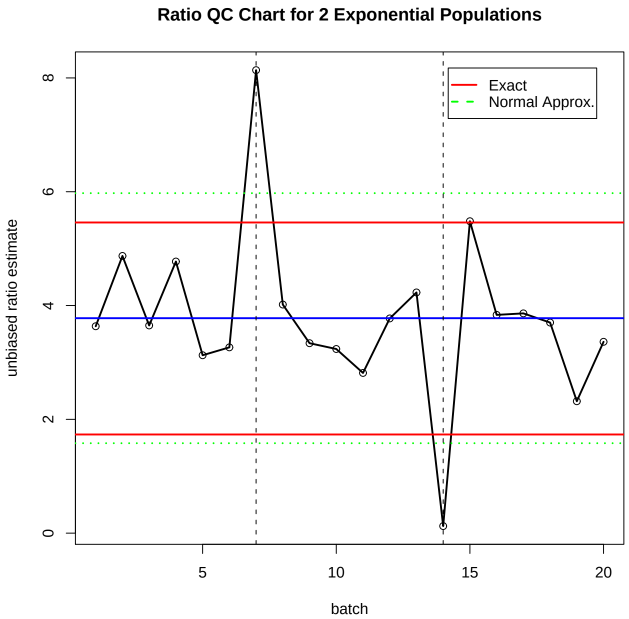

Figure 2: Exact vs. Normal approximation for Quality Control bounds: large sample

4 Conclusion

We derived the density of our pivot for the ratio of two exponential means. We then used theoretical graphs to check our work and then we solved for our upper and lower quantiles. It is safe to say that as we increase our sample sizes, we are able to draw more conclusions. The difficulty in this research was the lack of data for two process. We plan on extending our research and finding data for two processes instead of creating our own examples. We would also like to do research using Weibull distributions because the outcomes are more broad.

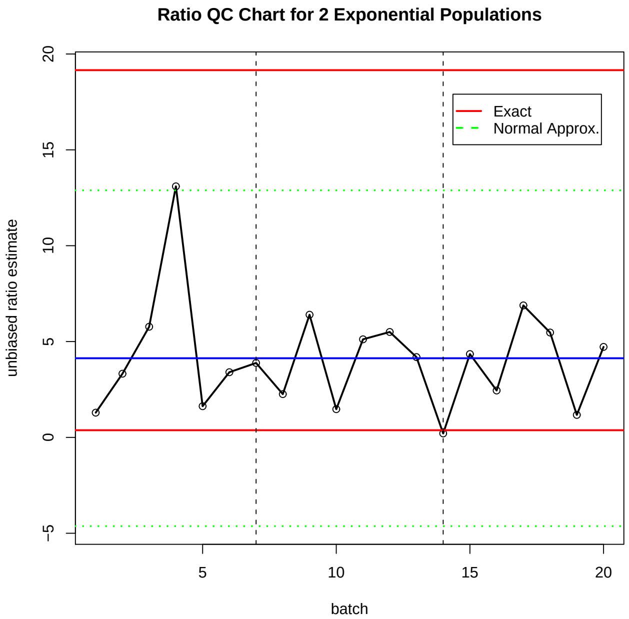

Figure 3: Exact vs. Normal approximation for Quality Control bounds: small sample

References

Hogg, R.V., and Tanis, E.A. (2006), *Probability and Statistical Inference 7e*, New Jersey: Prentice Hall.

Levine, D.M., Ramsey, P.P., and Smidt, R.K. (2001), *Applied Statistics for Engineers and Scientists*, New Jersey: Prentice Hall.

Appendix: Original method for finding Z critical points

Instead having to integrate and find the value for $Z_{\frac{\alpha}{2}}$ by hand, we decided to turn this into an actual Beta density.

$$\int_0^{Z_{\frac{\alpha}{2}}} \left(\frac{n_2 n_1}{(n_1 z + n_2)^2}\right)\left(\frac{1}{\beta(n_1, n_2)}\right) \times \left(\frac{n_2}{n_{1z} + n_2}\right)^{n_2 - 1}\left(1 - \frac{n_2}{n_1 z + n_2}\right)^{n_1 - 1} dz = 1 - \alpha$$

Let's use change of variables to reduce the integration. Let

$$\begin{aligned} x &= \frac{n_2}{n_1 z + n_2} = n_2(n_1 z + n_2)^{-1} \\ dx &= -\frac{n_1 n_2}{(n_1 z + n_2)^2} dz \end{aligned}$$

Since we changed the variables, we must now changes the bounds of the integration. That is, If $z = 0 \Rightarrow x = 1$, and if $z = z_{\frac{\alpha}{2}} \Rightarrow x = \frac{n_2}{n_1 z_{\frac{\alpha}{2}} + n_2}$. Therefore, the left hand side of the equation becomes

$$\int_1^{\frac{n_2}{n_1 + z_{\frac{\alpha}{2}} + n_2}} \frac{1}{\beta(n_1, n_2)} x^{n_2 - 1}(1 - x)^{n_1 - 1}(-dx)$$

which is an integral under a beta density. The negative sign in front of the dx allows us to switch the bounds. The equation now looks like:

$$\int_{\frac{n_2}{n_1 + z_{\frac{\alpha}{2}} + n_2}}^1 \frac{1}{\beta(n_1, n_2)} x^{n_2 - 1}(1 - x)^{n_1 - 1} dx = 1 - \frac{\alpha}{2}$$

For brevity, let $c_1 = \dfrac{n_2}{n_1 z_{\frac{\alpha}{2}+n_2}}$ such that

$$1 - \frac{\alpha}{2} = \int_c^1 (Beta(x : \alpha = n_2, \beta = n_1)dx)$$

So c_1 is the $\alpha/2$ percentile of a beta density, which can be easily be obtained using statistical software. In R statistical software we can use the qbeta() function such that:

$$c_1 \quad = \quad qbeta\left(\frac{\alpha}{2}, n_2, n_1\right) \tag{9}$$

In terms of c_1, $Z_{\alpha/2}$ is

$$\begin{aligned} n_1 z_{\frac{\alpha}{2}} + n_2 &= \frac{n_2}{c_1} \\ z_{\frac{\alpha}{2}} &= \frac{\frac{n_2}{c_1} - n_2}{n_1}. \end{aligned}$$

Similarly, we obtain c_2 from the beta density

$$c_2 = qbeta(1 - \frac{\alpha}{2}, n_2, n_1)$$

such that

$$z_{1-\frac{\alpha}{2}} = \frac{\frac{n_2}{c_2} - n_2}{n_1}.$$

Now that we have solved for $Z_{\frac{\alpha}{2}}$ and $1 - Z_{\frac{\alpha}{2}}$ we can now use them in our previous probability statement. Keep in mind that c_1 is our smaller value and c_2 is our larger value. This is because we want our Upper Bound to divide by the smaller value and the Lower Bound to divide by the larger value.