CONFIDENCE INTERVALS FOR VARIANCE COMPONENTS (Section 17.3.5)

In practice, these play the role for random effects that confidence intervals for contrasts play for fixed effects.

**Confidence intervals for $\sigma^2$:** These are constructed just as for fixed effects; see Section 3.4.6 or the class notes *Choosing Sample Sizes*.

**Confidence intervals involving $\sigma_T^2$:** Three types of confidence intervals are of interest: for $\sigma_T^2$, for $\sigma_T^2/\sigma^2$, and for $\sigma_T^2/(\sigma_T^2 + \sigma^2)$. The first cannot be done exactly, so we’ll take that last.

**Confidence intervals for $\sigma_T^2/\sigma^2$:** We use the fact (see notes *Testing for Treatment Effect as a Proportion of Error Variance*) that

$$\frac{\frac{\text{MST}}{(c\sigma_T^2 + \sigma^2)}}{\frac{\text{MSE}}{(\sigma^2)}} \sim F(v-1, n-v),$$

where $c$ is a certain constant defined in terms of $n,v$, and the $r_i$'s; $c = r$ if the design is balanced (See notes *Random Effects Models* or Section 17.3)

If we want a $(1-\alpha)100\%$ CI for $\sigma_T^2/\sigma^2$, take

- $f_1 = F(v-1, n-v, 1- \alpha/2)$ (so that there is area $\alpha/2$ to the left of $f_1$ in the $F(v-1, n-v)$ distribution), and
- $f_2 = F(v-1, n-v, \alpha/2)$ (so that there is area $\alpha/2$ to the right of $f_2$ in the $F(v-1, n-v)$ distribution). [Draw a picture!]

Then

$$\text{Prob} \left( f_1 \leq \frac{\text{MST}}{\frac{(c\sigma_T^2 + \sigma^2)}{\text{MSE}}} \leq f_2 \right) = 1 - \alpha,$$

or equivalently,

$$\text{Prob} \left( f_1 \leq \frac{[\text{MST}]}{[\text{MSE}]} \left[ \frac{\sigma^2}{(c\sigma_T^2 + \sigma^2)} \right] \leq f_2 \right) = 1 - \alpha,$$

The left inequality is equivalent to

$$\frac{(c\sigma_T^2 + \sigma^2)}{\sigma^2} \leq (\text{MST}/\text{MSE})(1/f_1),$$

or

$$c(\sigma_T^2/\sigma^2) + 1 \leq (\text{MST}/\text{MSE})(1/f_1),$$

which is equivalent to

$$c(\sigma_T^2/\sigma^2) \leq (\text{MST}/\text{MSE})(1/f_1) - 1$$

The right inequality is equivalent to
\[
\frac{\text{MST/MSE}(1/f_2)}{\sigma^2} \leq \frac{c\sigma_T^2 + \sigma^2}{\sigma^2} = c(\sigma_T^2/\sigma^2) + 1,
\]
which is equivalent to
\[
\frac{\text{MST/MSE}(1/f_2) - 1}{\sigma^2} \leq c(\sigma_T^2/\sigma^2)
\]
So
\[
\text{Prob} \left( \frac{\text{MST/MSE}(1/f_2) - 1}{\sigma^2} \leq \frac{\sigma_T^2}{\sigma^2} \leq \frac{\text{MST/MSE}(1/f_1) - 1}{\sigma^2} \right) = 1 - \alpha.
\]
Thus
\[
\left( \frac{\text{MST/MSE}(1/f_2)}{\sigma_T^2/\sigma^2} - 1 \right), \left( \frac{\text{MST/MSE}(1/f_1)}{\sigma_T^2/\sigma^2} - 1 \right)
\]
is the desired confidence interval. (This means ____________________________)

Note: Conceivably the left hand endpoint could be less than 0, which is unrealistic. If it is < 0, do not give in to the temptation to replace it by zero; that would give the false impression of a smaller confidence interval than warranted.

Example: Use the loom data to find a 95% confidence interval for \( \sigma_T^2/\sigma^2 \).

Confidence intervals for \( \sigma_T^2/(\sigma_T^2 + \sigma^2) \): the proportion of the total variance if the response attributable to the treatment level: Such confidence intervals are readily obtained from confidence intervals for \( \sigma_T^2/\sigma^2 \) as follows. Divide both numerator and denominator of \( \sigma_T^2/(\sigma_T^2 + \sigma^2) \) by \( \sigma^2 \) to obtain
\[
\frac{\sigma_T^2}{\sigma_T^2 + \sigma^2} = f\left( \frac{\sigma_T^2}{\sigma^2} \right), \text{ where } f(x) = x/(x + 1) = \frac{1}{1 + \frac{1}{x}}.
\]
From the last formula for \( f(x) \), we can see that \( f(x) \) is an increasing function of \( x \). Thus if (a,b) is a (1-\( \alpha \))100% confidence interval for \( \sigma_T^2/\sigma^2 \), then \( f(a), f(b) \) = (a/(a + 1), b/(b + 1)) is a (1-\( \alpha \))100% confidence interval for \( \sigma_T^2/(\sigma_T^2 + \sigma^2) \).

Note: \( \sigma_T^2/(\sigma_T^2 + \sigma^2) \) is sometimes called the “population intraclass correlation coefficient” (Caution: The phrase “intraclass correlation coefficient” is also used to refer to other things.)

Example: With the loom data, find a 95% confidence interval for \( \sigma_T^2/(\sigma_T^2 + \sigma^2) \).

Confidence intervals for \( \sigma_T^2 \): There is no exact method. There are several approximate methods. Here is one. It is useful if \( \sigma_T^2 \) is not too small, and is adaptable to more complicated models.

Recall that 
\[
U = (1/c)(\text{MST} - \text{MSE})
\]
is an unbiased estimator of \( \sigma_T^2 \). If we knew its distribution, we could use that to get confidence intervals for \( \sigma_T^2 \) in the usual way. However, \( U \) does not have a tractable distribution. But it is true that
\[
U/\sigma_T^2 = \chi^2(x)/x,
\]
where
\[ x = \frac{(msT - msE)^2}{(msT)^2 (v - 1) + (msE)^2 (n - v)} \]

(Note: This formula is given correctly on p. 605 of the text, but incorrectly on p. 600.)
x is not usually an integer, so we need to interpret degrees of freedom in the \( \chi^2 \) distribution as the parameter in a formula for the pdf. (This is analogous to the two-sample, unequal variance t-test.)

Thus (Draw a picture!)

\[ P( \chi^2(x, 1 - \alpha/2) < \frac{xU}{\sigma_T^2} < \chi^2(x, \alpha/2) ) \approx 1 - \alpha, \]

where \( < \) means “is less than or approximately equal to”, and \( \chi^2(x, \beta) \) is the value with proportion \( \beta \) of the \( \chi^2(x) \) distribution to its right.

The left and right approximate inequalities are, respectively, equivalent to

\[ \sigma_T^2 < \frac{xU}{\chi^2(x, 1 - \alpha/2)} \quad \text{and} \quad \sigma_T^2 > \frac{xU}{\chi^2(x, \alpha/2)}. \]

Thus if

\[ u = \frac{1}{c}(msT - msE) \] (which is our estimate for \( \sigma_T^2 \)), then

\((xu/\chi^2(x, \alpha/2), xu/\chi^2(x, 1 - \alpha/2))\) is an approximate \((1-\alpha)100\%\) confidence interval for \( \sigma_T^2 \).

**Example:** With the loom data, find a 95% confidence interval for \( \sigma_T^2 \).