

Activity with random effects for Lab 9, ST512

1. Recall the experiment with $t = 5$ sires (male beef cattle parents), and $n = 8$ offspring from each, leading to measured birthweights of $N = 40$ offspring. We considered a random effects model:

$$Y_{ij} = \mu + T_i + E_{ij}$$

where

- μ is an unknown population mean
- T_i is a random sample of sire effects from a population with unknown variance σ_T^2
- E_{ij} is a random sample of errors with unknown error variance σ^2 .

In this activity, we will simulate many datasets from this model to study the *sampling distributions* of variance statistics:

- (a) We'll see that $\hat{\sigma}^2$ and $\hat{\sigma}_T^2$ have non-normal sampling distributions. While these statistics are, on average, on target, they have large variability and can occasionally dramatically overshoot the variance components they are trying to estimate.
- (b) We'll see that if we compute a confidence interval for μ from a (misspecified) fixed effects model, we'll get the wrong answer. That is, the formula

$$\bar{y}_{+++} \pm t(.025, N - t) \sqrt{\frac{MS(E)}{N}}$$

will lead to a much lower coverage rate than 95%.

- (c) We'll see that the appropriate 95% confidence interval takes a different form:

$$\bar{y}_{+++} \pm t(.025, t - 1) \sqrt{\frac{MS(Trt)}{N}}$$

(Note that the degrees of freedom and mean square term are different.)

SAS concepts

- the SAS macro language. (Run the `datagen` macro with different values of the macro variables, like `nsim=10` and `nsim=20`)
- `by` processing (for each level of a variable such as `simulation`)
- the output delivery system (`ods`)
- `ods trace on;` to get info in the log about available datasets
- `ods listing close;` to stop sending output to the output window.