STA 6384, Report 1.3

Carson Slater *Baylor University*

Problem: Work problem 1.3, p. 29 of Agresti.

An experiment studies the number of insects that survive a certain dose of an insecticide, using several batches of insects of size n each. The insects are sensitive to factors that vary among batches during the experiment but were not measured, such as temperature level. Explain why the distribution of the number of insects per batch surviving the experiment might show overdispersion relative to a bin(n, π) distribution.

Solution to 1.3. The binomial distribution Bin(n, π) assumes that each of the n trials (insects) within a batch is:

- independent of the others,
- has the same probability π of success (i.e., survival),
- and that this probability π is constant across batches.

However, in this experiment, there are unmeasured factors (e.g., temperature, humidity, genetic variability, or handling differences) that vary between batches and affect the survival probability. As a result, the probability of survival π is not constant, but rather varies from batch to batch.

Why overdispersion occurs: If π varies across batches, then the variability in the number of surviving insects across batches exceeds the binomial variance. For a binomial distribution, the variance is:

$$\operatorname{Var}(Y) = n\pi(1-\pi)$$

But if π is itself a random variable (e.g., due to batch-level variation), then the law of total variance gives:

$$\operatorname{Var}(Y) = \mathbb{E}[n\pi(1-\pi)] + n^2 \operatorname{Var}(\pi)$$

The second term, $n^2 \operatorname{Var}(\pi)$, inflates the variance beyond the binomial variance. This is known as *overdispersion.*

In conclusion, the observed number of surviving insects per batch may show overdispersion relative to a Bin(n, π) distribution because the true survival probability varies across batches due to unmeasured heterogeneity. This leads to greater variability in the number of survivors than the binomial model would predict. [ChatGPT 4o–assisted]