STA 6384, Report 3.3

Carson Slater *Baylor University*

Problem: Work problem 3.25, pp. 108-109 in Agresti.

3.25 For comparing two binomial samples with fixed sample sizes, show that the standard error of a log odds ratio increases when, for either sample, the absolute difference of proportions of successes and failures increases. (Hint: Show that the asymptotic variance is minimized when each binomial probability is 0.50. In particular, when an outcome is relatively uncommon, estimates of the log odds ratio tend to be imprecise.)

Let's consider two independent binomial samples with sizes n_1 and n_2. Let the probabilities of success in the two populations be π_1 and π_2, respectively. The sample proportions of successes are $\hat{\pi}_1$ and $\hat{\pi}_2$.

The odds ratio is defined as $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$. The natural logarithm of the sample odds ratio, $\log(\hat{\theta})$, is a key statistic for inference.

Asymptotic Variance of the Log Odds Ratio

The asymptotic variance of the log odds ratio estimator is given by:

$$V(\log(\hat{\theta})) = \frac{1}{n_1\pi_1(1-\pi_1)} + \frac{1}{n_2\pi_2(1-\pi_2)}$$

The standard error (SE) is the square root of the estimated variance, where we substitute the sample proportions $\hat{\pi}_1$ and $\hat{\pi}_2$ for the population probabilities π_1 and π_2. To show that the SE increases under a certain condition, it is sufficient to show that the variance increases.

Minimizing the Variance

The total variance is a sum of two terms, one for each sample. Let's analyze the contribution from a single sample, which is of the form:

$$V_i = \frac{1}{n_i\pi_i(1-\pi_i)}$$

To find when this term is minimized, we need to find when its denominator, $n_i\pi_i(1-\pi_i)$, is maximized. For a fixed sample size n_i, we only need to maximize the function $f(\pi) = \pi(1-\pi) = \pi - \pi^2$.

We can find the maximum of $f(\pi)$ for $\pi \in (0,1)$ using calculus. The first derivative is:

$$f'(\pi) = \frac{d}{d\pi}(\pi - \pi^2) = 1 - 2\pi$$

Setting the derivative to zero, $1 - 2\pi = 0$, gives $\pi = 0.50$. To confirm this is a maximum, we check the second derivative:

$$f''(\pi) = -2$$

Since the second derivative is negative, $f(\pi) = \pi(1-\pi)$ has a global maximum at $\pi = 0.50$. This means the variance contribution V_i is **minimized** when the probability of success is exactly 0.50.

Connecting to the Absolute Difference

The problem asks about the "absolute difference of proportions of successes and failures." For a given sample, this is $|\pi - (1-\pi)| = |2\pi - 1|$.

- When $\pi = 0.50$, the absolute difference is $|2(0.5) - 1| = 0$. This is the point where the variance is **minimized**.
- As π moves away from 0.50 towards either 0 or 1, the value of $|2\pi - 1|$ increases.
- As π moves away from its maximum at 0.50, the value of the function $f(\pi) = \pi(1-\pi)$ decreases.

Since the variance term $V_i = \frac{1}{n_i f(\pi_i)}$ is inversely proportional to $f(\pi_i)$, a decrease in $f(\pi_i)$ leads to an increase in V_i.

Therefore, as the absolute difference $|2\pi_i - 1|$ increases, the variance contribution V_i from that sample also increases.

Conclusion

The total variance, $V = V_1 + V_2$, is the sum of the individual variance components. If the absolute difference $|\pi_i - (1-\pi_i)|$ increases for either sample $i = 1$ or $i = 2$, the corresponding variance term V_i increases. Since all terms are positive, this causes the total variance V to increase. Consequently, the standard error, $\mathrm{SE}(\log(\hat{\theta})) = \sqrt{V}$, also increases.

This confirms the hint's final point: when an outcome is relatively uncommon, its probability π is close to 0 or 1. This is far from the variance-minimizing value of 0.50, leading to a large standard error and thus, imprecise estimates of the log odds ratio.