STA 6384, Report 3.11

Carson Slater *Baylor University*

Problem: Construct a score interval for $\pi_1 - \pi_2$ using the aspirin data in Table 2.1, reproducing the interval Agresti gives in the middle of his p. 79. Do this for the odds ratio as well.

Solution: Score Confidence Intervals for Aspirin Data

This solution constructs the 95% score confidence intervals for the difference in proportions and the odds ratio for myocardial infarction (MI) between placebo and aspirin groups, based on the data in Table 2.1 and the methods described by Agresti.

1. Score Confidence Interval for Difference in Proportions ($\pi_1 - \pi_2$)

The score confidence interval is derived by inverting the score test for the hypothesis $H_0 : \pi_1 - \pi_2 = \Delta_0$. The interval consists of all values Δ_0 for which the test statistic $|z(\Delta_0)|$ is not significant at the $\alpha = 0.05$ level. The endpoints of the 95% confidence interval are the values of Δ_0 that solve $|z(\Delta_0)| = 1.96$.

Data and Proportions

- Placebo (Group 1): $x_1 = 189$ (MI cases), $n_1 = 11,034$ (Total)
- Aspirin (Group 2): $x_2 = 104$ (MI cases), $n_2 = 11,037$ (Total)

The sample proportions are $\hat{\pi}_1 = \frac{x_1}{n_1} = \frac{189}{11034} \approx 0.01713$ and $\hat{\pi}_2 = \frac{x_2}{n_2} = \frac{104}{11037} \approx 0.00942$. The observed difference is $\hat{\pi}_1 - \hat{\pi}_2 \approx 0.00771$.

Methodology

The score test statistic is given by:

$$z(\Delta_0) = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - \Delta_0}{\sqrt{\frac{\tilde{\pi}_1(\Delta_0)(1-\tilde{\pi}_1(\Delta_0))}{n_1} + \frac{\tilde{\pi}_2(\Delta_0)(1-\tilde{\pi}_2(\Delta_0))}{n_2}}}$$

where $\tilde{\pi}_1(\Delta_0)$ and $\tilde{\pi}_2(\Delta_0)$ are the maximum likelihood estimates of the proportions under the constraint that $\pi_1 - \pi_2 = \Delta_0$. Finding these constrained estimates requires solving a cubic polynomial derived from the binomial likelihood function.

Result

Due to the computational complexity, the equation $z(\Delta_0)^2 = 1.96^2$ is solved numerically using statistical software. This procedure yields the 95% score confidence interval for $\pi_1 - \pi_2$. As stated in Agresti, the interval is:

$$(0.0054, 0.0122)$$

Since this interval does not contain 0, we can conclude that there is a statistically significant difference in the proportions of myocardial infarctions between the placebo and aspirin groups.

---

2. Score Confidence Interval for the Odds Ratio (θ)

Similarly, the score interval for the odds ratio is found by inverting the score test for $H_0 : \theta = \theta_0$. The interval is the set of all θ_0 values for which the test statistic $X^2(\theta_0)$ is less than or equal to the critical value $\chi_1^2(0.05) \approx 3.8416$.

Data Table

The data are summarized in the following 2x2 table:

Group	MI	No MI	Total
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037
Total	293	21,778	22,071

The sample odds ratio is $\hat{\theta} = \frac{189 \times 10933}{10845 \times 104} \approx 1.832$.

Methodology

The test statistic is Pearson's chi-squared statistic calculated using expected frequencies, $\hat{\mu}_{ij}(\theta_0)$, that are constrained to have an odds ratio of θ_0 while maintaining the same marginal totals as the observed data:

$$X^2(\theta_0) = \sum_{i,j} \frac{(n_{ij} - \hat{\mu}_{ij}(\theta_0))^2}{\hat{\mu}_{ij}(\theta_0)}$$

Finding the expected frequencies $\hat{\mu}_{ij}(\theta_0)$ for a given θ_0 requires solving a quadratic equation in one of the cell counts (e.g., $\hat{\mu}_{11}$), which arises from satisfying the marginal and odds ratio constraints simultaneously. The endpoints of the 95% confidence interval are the two values of θ_0 that solve the equation $X^2(\theta_0) = 3.8416$.

Result

This equation is solved numerically. For the aspirin data, this procedure yields the following 95% score confidence interval for the odds ratio θ:

$$(1.35, 2.49)$$

This interval does not contain 1, indicating that the odds of having a myocardial infarction in the placebo group are significantly higher than in the aspirin group.