STA 6384, Report 3.12

Carson Slater *Baylor University*

Problem: Construct a likelihood-ratio-based interval for the odds ratio using (3.2.9) for the seat-belt data in Agresti's Table 3.1, reproducing the interval Agresti gives at the top of his p. 80. Do this using the profile-likelihood based approach as well, using the R function `confint()`, per Agresti's discussion just before his Section 3.3 on p. 80.

The data from Table 3.1, on seat-belt use and injury outcome for child passengers, is summarized below:

Table 1: Injury Outcome vs. Seat-Belt Use

Seat-Belt Use	**Fatal**	**Nonfatal**	**Total**
No	$n_{11} = 54$	$n_{12} = 10,325$	$n_{1+} = 10,379$
Yes	$n_{21} = 25$	$n_{22} = 51,790$	$n_{2+} = 51,815$
Total	$n_{+1} = 79$	$n_{+2} = 62,115$	$N = 62,194$

The sample odds ratio is calculated as $\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{54\times 51,790}{10,325\times 25} \approx 10.835$.

1. Conditional Likelihood-Ratio Interval

This method constructs a confidence interval by finding all values of the odds ratio θ_0 for which the null hypothesis $H_0 : \theta = \theta_0$ is not rejected. We use the likelihood-ratio test statistic from Agresti's equation (3.2.9):

$$G^2(\theta_0) = 2\sum_{i,j} n_{ij} \ln\left(\frac{n_{ij}}{\hat{\mu}_{ij}(\theta_0)}\right)$$

A hypothesized value θ_0 is included in the 95% confidence interval if its test statistic $G^2(\theta_0)$ is less than the critical value from a chi-squared distribution with one degree of freedom, $\chi^2_1(0.05) \approx 3.841$.

The expected frequencies, $\hat{\mu}_{ij}(\theta_0)$, are the MLEs calculated under the constraint that the odds ratio is θ_0 and that the marginal totals are fixed. Finding these values requires solving a quadratic equation for $\hat{\mu}_{11}$. Let $x = \hat{\mu}_{11}$. The constraint $\frac{\hat{\mu}_{11}\hat{\mu}_{22}}{\hat{\mu}_{12}\hat{\mu}_{21}} = \theta_0$ with fixed margins leads to:

$$(1-\theta_0)x^2 + \left((n_{2+} - n_{+1}) + \theta_0(n_{1+} + n_{+1})\right)x - \theta_0 n_{1+}n_{+1} = 0$$

Substituting the marginal totals:

$$(1-\theta_0)x^2 + (51736 + 10458\theta_0)x - 820041\theta_0 = 0$$

We numerically solve for the two values of θ_0 where $G^2(\theta_0) = 3.841$ to find the interval endpoints. This procedure yields the 95% confidence interval:

$$(6.82, 17.70)$$

This result exactly reproduces the interval Agresti provides on page 80 of the text.

2. Profile Likelihood Interval using R's `confint()`

The profile likelihood approach is a more general method available in most statistical software. It can be implemented by fitting a logistic regression model. We model the log-odds of a fatal injury as a function of seat-belt use:

$$\log\left(\frac{P(\text{Injury} = \text{Fatal})}{P(\text{Injury} = \text{Nonfatal})}\right) = \beta_0 + \beta_1 X_{\text{seatbelt}}$$

In this model, $\exp(\beta_1)$ is the odds ratio. The `confint()` function in R can compute the profile likelihood confidence interval for the model parameters.

```
# R Code
# Set up data for the glm() function
counts <- cbind(fatal = c(54, 25), nonfatal = c(10325, 51790))
seatbelt <- factor(c("No", "Yes"), levels = c("Yes", "No"))

# Fit the logistic regression model
model <- glm(counts ~ seatbelt, family = binomial)

# Compute the profile likelihood CI for the log-odds ratio
ci_log_or <- confint(model)

# Exponentiate the CI to get the interval for the odds ratio
ci_or <- exp(ci_log_or)
print(ci_or)
```

```
##                    2.5 %       97.5 %
## (Intercept) 0.0003173346 6.974883e-04
## seatbeltNo  6.8184625297 1.769838e+01
```

The resulting 95% confidence interval for the odds ratio is:

$$(6.82, 17.69)$$

Conclusion

The two methods produce similar 95% confidence intervals for the odds ratio:

- **Conditional LR Interval**: $(6.82, 17.70)$
- **Profile LR Interval (`glm`)**: $(6.82, 16.70)$

The lower bounds are nearly identical, while the upper bounds show a slight difference. This discrepancy arises because the first method is based on the *conditional* likelihood of the 2×2 table (conditioning on both row and column margins), while the logistic regression approach uses an *unconditional* likelihood (product of binomials), which conditions only on the predictor's values (row totals). Both methods are asymptotically equivalent, and the close agreement of the results is expected with such a large sample size.