STA 6384, Report 3.1

Carson Slater *Baylor University*

Problem: Use R to reproduce results in the seat-belt use data (Section 3.1.2, p. 70).

Data Setup

First, we construct the 2x2 contingency table from the source (Table 3.1). The table displays counts of fatal and nonfatal outcomes for children under 18, categorized by whether they were wearing a seat belt.

Table 1: Table 1: Injury Outcome by Seat-Belt Use

	Fatal	Nonfatal
No	54	10325
Yes	25	51790

Odds Ratio (θ)

The odds of a fatal outcome for a child not wearing a seat belt are compared to the odds of a fatal outcome for a child wearing one. The sample **odds ratio ($\hat{\theta}$)** is calculated as:

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

The calculated sample odds ratio is **10.83**, which matches the value of **10.83** cited in the text. This indicates that the odds of a fatal injury were nearly 11 times higher for children who were not wearing a seat belt compared to those who were.

Confidence Interval for the Odds Ratio

To determine the precision of this estimate, we calculate a 95% confidence interval. This is done by first finding the confidence interval for the **log odds ratio** and then exponentiating the endpoints.

1. **Log Odds Ratio and Standard Error**: The natural logarithm of the odds ratio and its standard error are calculated as follows:

$$\log(\hat{\theta})$$
$$\hat{\sigma}(\log\hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

The resulting log odds ratio is **2.383** and its standard error is **0.242**. These values match the **2.383** and **0.242** reported in the text.

2. **95% Confidence Interval for Log Odds Ratio**: The 95% confidence interval is found using the formula $\log(\hat{\theta}) \pm 1.96 \times \hat{\sigma}(\log \hat{\theta})$.

   The 95% confidence interval for the log odds ratio is **(1.908, 2.857)**. This aligns with the interval **(1.908, 2.857)** from the text.

3. **95% Confidence Interval for Odds Ratio**: Finally, we exponentiate the lower and upper bounds of the log odds ratio's confidence interval to transform it back to the odds ratio scale.

   The 95% confidence interval for the true odds ratio is **(6.74, 17.42).** This result successfully reproduces the interval of **(6.74, 17.42)** reported in the text. The interval does not include 1, confirming a statistically significant and strong association between not wearing a seat belt and an increased odds of a fatal injury.

Code Appendix

```r
# --- R Code for Reproduction ---

# 1. Create the 2x2 contingency table from the data
seatbelt_data <- matrix(c(54, 10325, 25, 51790), nrow = 2, byrow = TRUE,
                        dimnames = list(SeatBelt = c("No", "Yes"),
                                        Outcome = c("Fatal", "Nonfatal")))

# Display the table
# print(seatbelt_data)
# knitr::kable(seatbelt_data)


# 2. Extract individual cell counts from the matrix
n11 <- seatbelt_data[1, 1] # Outcome: Fatal, SeatBelt: No
n12 <- seatbelt_data[1, 2] # Outcome: Nonfatal, SeatBelt: No
n21 <- seatbelt_data[2, 1] # Outcome: Fatal, SeatBelt: Yes
n22 <- seatbelt_data[2, 2] # Outcome: Nonfatal, SeatBelt: Yes


# 3. Calculate the sample odds ratio (theta-hat)
# Formula: (n11 * n22) / (n12 * n21)
odds_ratio <- (n11 * n22) / (n12 * n21)
# print(paste("Sample Odds Ratio:", round(odds_ratio, 2)))


# 4. Calculate the natural log of the odds ratio
```

```r
log_odds_ratio <- log(odds_ratio)
# print(paste("Log Odds Ratio:", round(log_odds_ratio, 3)))


# 5. Calculate the standard error of the log odds ratio
# Formula: sqrt(1/n11 + 1/n12 + 1/n21 + 1/n22)
se_log_or <- sqrt(1/n11 + 1/n12 + 1/n21 + 1/n22)
# print(paste("Standard Error of Log Odds Ratio:", round(se_log_or, 3)))


# 6. Calculate the 95% confidence interval for the log odds ratio
# Formula: log(OR) +/- 1.96 * SE(log(OR))
z_value <- 1.96
log_ci_lower <- log_odds_ratio - z_value * se_log_or
log_ci_upper <- log_odds_ratio + z_value * se_log_or
# print(paste0("95% CI for Log Odds Ratio: (", round(log_ci_lower, 3), ", ", r


# 7. Exponentiate the log CI to get the 95% CI for the odds ratio
ci_lower <- exp(log_ci_lower)
ci_upper <- exp(log_ci_upper)
# print(paste0("95% CI for Odds Ratio: (", round(ci_lower, 2), ", ", round(ci_
```