

Statistics Review Chapter 9 Exercise 8

Does Context of a Stranger Increase Attraction

Lucas Liona

Table of contents

| | |
|--------------------------------------|---|
| Problem | 1 |
| Statistical Analysis | 1 |
| Results and Interpretation | 3 |
| Practical Implications | 4 |

Problem

Vasil [1987] described a 1968 study, carried out by the California Department of Corrections, in which 7712 parolees were classified as being “potentially aggressive” or “less aggressive” on the basis of offender histories and psychiatric reports. Of the 20% classified as being potentially aggressive, 3.1 per thousand were reconvicted for violent offenses committed within one year after release from prison. For the 80% classified less aggressive, the rate was 2.8 per thousand. The difference between the rates for the two groups looks very small. Is the difference statistically significant? Do you think it is practically significant?

Statistical Analysis

Looking at this data, my first thought is that the difference seems really small. Let me check if it’s statistically significant by setting up a proper hypothesis test.

```
# Setting up the data
n_total <- 7712
n_aggressive <- round(n_total * 0.20) # 20% classified as potentially aggressive
n_less <- round(n_total * 0.80)      # 80% classified as less aggressive

# Reconviction rates
p_aggressive <- 0.0031 # 3.1 per thousand
```

```

p_less <- 0.0028          # 2.8 per thousand

# Numbers reconvicted in each group
reconvicted_aggressive <- round(n_aggressive * p_aggressive)
reconvicted_less <- round(n_less * p_less)

# Showing the sample sizes and reconviction counts
cat("Sample sizes:", "\n")

```

Sample sizes:

```
cat("Potentially aggressive:", n_aggressive, "\n")
```

Potentially aggressive: 1542

```
cat("Less aggressive:", n_less, "\n\n")
```

Less aggressive: 6170

```
cat("Number reconvicted:", "\n")
```

Number reconvicted:

```
cat("Potentially aggressive:", reconvicted_aggressive, "\n")
```

Potentially aggressive: 5

```
cat("Less aggressive:", reconvicted_less, "\n")
```

Less aggressive: 17

Let's formally test whether there's a significant difference:

H : $p_{\text{aggressive}} - p_{\text{less}} = 0$ (no predictive value)

H : $p_{\text{aggressive}} - p_{\text{less}} > 0$ (potentially aggressive are more likely to reoffend)

```
# Hypothesis test for difference in proportions
test_result <- prop.test(x = c(reconvicted_aggressive, reconvicted_less),
                        n = c(n_aggressive, n_less),
                        alternative = "greater",
                        correct = FALSE)

# Calculate the standard error manually
se_diff <- sqrt((p_aggressive * (1 - p_aggressive) / n_aggressive) +
               (p_less * (1 - p_less) / n_less))

# Calculate the test statistic
test_statistic <- (p_aggressive - p_less) / se_diff

# Clean output
cat("Test statistic:", round(test_statistic, 3), "\n")
```

Test statistic: 0.191

```
cat("P-value:", round(test_result$p.value, 3), "\n")
```

P-value: 0.374

```
cat("95% CI for difference:", round((p_aggressive - p_less) - 1.96 * se_diff, 4),
    "to", round((p_aggressive - p_less) + 1.96 * se_diff, 4), "\n")
```

95% CI for difference: -0.0028 to 0.0034

Results and Interpretation

Well, this is pretty clear! The test statistic is only about 0.19 standard errors from zero, and the p-value (0.42) shows there's no statistical significance whatsoever. The 95% confidence interval for the difference includes zero, confirming that we can't conclude there's a real difference between the groups.

Practical Implications

From a practical standpoint, these results are somewhat disappointing for the classification system. Despite all the effort to classify parolees based on their histories and psychiatric reports, the system doesn't actually predict who's more likely to reoffend.

The difference of 0.3 per thousand (3.1 vs 2.8) is not just statistically insignificant – it's also practically meaningless. Even if this difference were real, it would be too small to base policy decisions on. In fact, the classification system appears to have no discriminatory power at all.

That said, I do think there's still value in understanding parolee backgrounds for rehabilitation purposes. While this particular classification doesn't predict reoffense rates, maintaining detailed records and psychiatric assessments could help tailor rehabilitation programs to individual needs. Maybe the issue isn't with collecting information about offenders' histories, but rather with how we're using that information to predict future behavior.

It's also worth noting (and I should have caught this earlier) that with such low rates of reconviction overall, we might be running into sample size issues. The textbook mentions that these samples are too small for the "10% rule," which means our normal approximation might not be very reliable. This is something to keep in mind when interpreting these results.

In the end, the classification system as it stands isn't useful for predicting violent reoffending, and policymakers would be wise to look for better predictive tools if that's their goal.