



(i) Fit a logistic model

Understanding the Problem:

We have a case-control study with:

- **Cases:** 1000 males with lung cancer.
- **Controls:** 1000 males without lung cancer.

The exposure variable is smoking:

- Among cases: 452 smokers, 548 non-smokers.
- Among controls: 215 smokers, 785 non-smokers.

Logistic Regression Model:

In logistic regression, we model the log-odds of the outcome (having lung cancer) as a linear function of the predictor (smoking).

Let:

- $Y = 1$ for cases (lung cancer), $Y = 0$ for controls.
- $X = 1$ for smokers, $X = 0$ for non-smokers.

The logistic model is:

$$\log \left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)} \right) = \beta_0 + \beta_1 X$$

Estimating Parameters:

In a case-control study, we cannot directly estimate $P(Y = 1|X)$ because the proportion of cases and controls is fixed by design. However, we can estimate the odds ratio, which is the exponentiated coefficient β_1 .

The odds ratio (OR) can be calculated from the 2x2 table:



Message DeepSeek



DeepThink (R1)



Search

