7.8 Practical 7

Datasets required: `oesophageal_data-1.dta` and `oesophageal_data-2.dta`

Introduction

We will use data from a case-control study of oesophageal cancer. Cases are 200 men diagnosed with oesophageal cancer in the Ille-et-Vilaine area of France between January 1972 and April 1974. Controls are 776 men who were sampled from an electoral register. The data are used extensively in the book Breslow, N.E. and Day, N.E. 1980. Statistical Methods in Cancer Research: Volume 1 The Analysis of Case-Control Studies. This book is freely available online at

`https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Statistical-Methods-In-Cancer-Research-Volume-I-The-Analysis-Of-Case-Control-Studies-1980`

The first dataset is called oesophageal_data1.dta and consists of one row per participant, and four variables as below:

Variable	Description
`case`	0=control, 1=case
`age_group`	age group 1: 25-34, 2: 35-44, 3: 45-54, 4: 55-64, 5: 65-74, 6: 75+
`tobacco_group`	tobacco intake group (in grams per day) 0: None, 1: 1-9, 2: 10-19, 3: 20-29, 4: 30+
`alcohol_grp`	alcohol intake group (in grams per day) 0: 0-39, 1: 40-79, 2: 80-119, 3: 120+

Aims

We will examine the association between oesophageal cancer as the outcome and a dichotomous smoking status variable (smoker or non-smoker) as an exposure, and attempt to answer the following questions:

1 Is smoking associated with the risk ofoesophageal cancer?

2 Does alcohol intake or age confound this relationship?

3 What is the best way to adjust for these potential confounders: as categorical or continuous variables?

We will also swap the exposure and outcome variables to explore what happens to the estimate of the odds ratio (and its confidence interval).

Analysis

1 Generate a new binary variable for tobacco intake which takes value 0 if intake is 0 and value 1 otherwise. Call this tobacco_2. Explore the associations between:

(a) binary tobacco status and alcohol group

(b) binary tobacco status and age

(c) alcohol and age

Discuss: Are there any associations which may affect your analysis?

2 (a) Use the two-by-two table for the binary tobacco variable and the case-control status to estimate the odds ratio. Use Woolf's method to obtain a 95% confidence interval for this estimate.

(b) Interpret your results.

(c) Compare your estimates to those obtained from the `mhodds` command.

3 Write down the algebraic form of the following logistic regression models, and then use the `glm` command in Stata to fit the models:

(a) with oesophageal cancer as the outcome and tobacco intake as the explanatory variable

(b) with tobacco intake as the outcome and oesophageal cancer as the explanatory variable

Discuss: Compare and contrast the results from the two analyses

4 We will now control for alcohol intake. Using the Mantel-Haenszel method `mhodds`, find the conditional odds ratio:

(a) with oesophageal cancer as the outcome and tobacco intake as the explanatory variable

(b) with tobacco intake as the outcome and oesophageal cancer as the explanatory variable

5 Repeat Q4 a) and b) using the `glm` command in Stata.

Discuss: Compare and contrast the results from the two models, and the two mhodds commands. What do the intercept parameters represent in each logistic model?

6 We will now also condition on age. Add the age group variable into the model with case-control status as the outcome and interpret the results.

Discuss: What impact does adding age to the model have on the odds ratio for tobacco status? What are two possible reasons for any impact that you see?

7 Age and alcohol intake are in fact available as continuous variables. These can be found in the `oesophageal_data-2.dta` dataset.

Fit a logistic model for oesophageal cancer with three explanatory variables: binary tobacco intake, continuous alcohol intake and continuous age.

Discuss: Compare the results from this model to your results from Q6. What are the advantages and disadvantages of the two approaches?

8 Present your results in a table or tables suitable for use in a paper or report.