

UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS
BIOSTATISTICS (MATH11230)

Assignment 1

- To be uploaded to Learn (Gradescope) 11. March, Monday 16:00, 2024.
 - This assignment is worth 5% of your final grade for the course.
 - Assignments should be typed (L^AT_EX, word, etc.).
 - Answers to questions should be in full sentences and should provide all necessary details.
 - Any output (e.g., graphs, tables) from R that you use to answer questions must be included with the assignment. Please submit your solution as pdf file in the gradescope submission box and submit your R code (.Rmd or .R file) in the learn submission box. Your R code won't be marked but we may look at it whilst marking.)
 - The assignment is out of 100 marks.
1. For each of the following scenarios, state, justifying, which measure (incidence or prevalence) should be calculated.
 - (a) **(2.5 marks)** Number of campers who developed gastroenteritis within a few days after eating potato salad at the dining hall.
 - (b) **(2.5 marks)** Number of persons who reported having diabetes as part of the National Health Interview Survey.
 - (c) **(2.5 marks)** Occurrence of acute myocardial infarction (heart attack) among participants during the first 10 years of follow-up of the Framingham Study?
 - (d) **(2.5 marks)** Number of children who have immunity to measles, either because they had the disease or because they received the vaccine.
 2. **(10 marks)** A randomised controlled trial was set up to assess the effects of two therapeutic regimens as preventative therapies for hip fractures: (i) medical therapy with a bisphosphonate versus (ii) training in weight-bearing physical exercise. Two hundred high-risk elderly women were selected for inclusion in the study. The women were allocated equally to the two therapeutic regimens. The outcome was whether the women experienced a fracture after 52 weeks. The outcome data are shown below:

Therapeutic regimen	Number of fractures
Bisphosphonate	22
Physical exercise	11

Perform an appropriate statistical analysis to assess the effects of physical exercise compared to bisphosphonate on fracture within 52 weeks.

3. **(10 marks)** Anecdotal evidence suggests that anxiety disorder may contribute to the irritable bowel syndrome (IBS), a chronic condition that affects the digestive system and whose main symptoms include nausea, alternating constipation and diarrhea, and no identifiable gastrointestinal pathology. Researchers administered an online questionnaire regarding IBS symptoms to 10 000 individuals, in the UK, who have an established diagnosis of anxiety disorder. The same questionnaire was administered to another 10 000 individuals, in the UK, but who do not have a diagnosis of anxiety disorder. The results are presented in the table below. Which of the following is true?

	IBS symptoms	No IBS symptoms
Anxiety disorder	4000	6000
No anxiety disorder	1000	9000

- (a) The incidence proportion of IBS symptoms among individuals with anxiety disorder is 40%.
- (b) The incidence rate of IBS symptoms among people with anxiety disorder is 40%.
- (c) The prevalence of IBS symptoms among individuals with anxiety disorder is 40%.
4. A case-control study of bladder cancer was conducted to test its putative association with dietary consumption of substance E . Cases and controls were selected randomly from male hospital patients between 50 and 60 years of age residing in a large metropolitan area. The tables below summarise the data by history of cigarette smoking. From previous studies, smoking is a known risk factor for bladder cancer. We shall note that consumption of substance E has been dichotomised into consumers E and non-consumers (not E). In addition, below, D denotes the occurrence of bladder cancer.

Table 1: Smokers		
	D	not D
E	35	20
not E	5	10

Table 2: Non-smokers		
	D	not D
E	10	20
not E	15	30

In what follows, assume that the potential confounding effect of age has been controlled by restriction.

- (a) **(5 marks)** What is the estimated unadjusted (with respect to smoking) odds ratio comparing cases and controls with regard to previous consumption of substance E ? Provide a 95% confidence interval as well.
- (b) **(5 marks)** Compare the two estimated smoking-specific odds ratios and their confidence intervals to each other and to the unadjusted odds ratio estimate. What do these results indicate?
- (c) **(5 marks)** Based on all the information that you have, is a single summary odds ratio appropriate? If yes, calculate it; if not, what is the best approach to summarise the data?
5. **(25 marks)** A case-control study investigated the role of several reproductive risk factors for breast cancer. A subset of the data relating to age and prior number of births is given in the table below.

Age	Prior number of births	Cases	Controls
< 40 years	0–1	24	58
	≥ 2	96	160
≥ 40 years	0–1	127	172
	≥ 2	353	718

Suppose that the prior number of births is given by the variable X_1 , coded such that $X_1 = 1$ if the prior number of births is 0–1, and $X_1 = 0$ if the prior number of births is ≥ 2 . Similarly, age is given by variable X_2 and coded such that $X_2 = 1$ if age is ≥ 40 years, and $X_2 = 0$ if age is < 40 years. Suppose the logistic model

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 X_2),$$

is fitted to the data. Interpret and estimate β_1 , β_2 , and β_3 . Here you are required to estimate each parameter manually and not to use the `glm` function (unless to check your answer, of course).

6. **(30 marks)** The data in the table below comes from a study about infant respiratory disease. Each cell of the table shows the number out of so many observed children who developed bronchitis or pneumonia in their first year of life, classified by sex and type of feeding.

Sex	Bottle only	Breast + supplement	Breast only
Boys	77/458	19/147	47/494
Girls	48/384	16/127	31/464

The major question of interest is whether the risk of illness is affected by the type of feeding. Also, is the risk the same for both sexes and, if there are differences between the feeding groups, are they the same for boys and girls? Use logistic regression models to answer these questions, clearly justifying and summarising your findings.

Note: The file `data_exbottle.xls` contains the data presented in the table above, where sex is coded as 1 for boys, and 0 for girls, and feeding type is coded as 1 for bottle only, 2 for breast plus supplement, and 3 for breast only. These variables should be coded as factors (i.e., through the use of indicator variables).