

EPM202: Statistical Methods in Epidemiology Assessed Assignment 2019/20

Contents

Aim of the assignment	1
Description of the dataset	1
Variables in the dataset	2
Research questions	2
Guidelines on how to write your report	3
General guidelines and advice	4
Criteria for grading	5
Submission format	6
Submission deadline	6

Aim of the assignment

The aim of this assignment is to test your understanding of the course material and your use of Stata. The assignment will give you an opportunity to put into practice the concepts and methods that you have learnt in the course.

It is advisable for you to undertake this assignment after you have studied up to and including Session 9, plus Session 12, of the CAL material. There is no single correct answer to the exercise. However, the tutors marking your assignment will look for an appropriate analysis leading to correct conclusions presented in a format along the lines of that used for a scientific paper (more detailed guidance on this is given below).

Description of the dataset

The data for this assignment come from a population-based study of female blood donors in the UK and is in the file named 'blood.dta' which can be found in the Assessed Assignment section in Moodle along with these instructions. The data were collected from 1,885 women attending a blood donation centre during 2015. Information on age, smoking, diet (red meat, tea and fruit consumption), iron supplementation and menopause status were collected via questionnaire. Additionally, height and weight was measured by a nurse and used to calculate BMI. In the UK, women must pass a test measuring blood haemoglobin ($>125\text{g/dl}$) before they are allowed to donate blood. In this study, 225 women failed the test and so were not allowed to donate blood on that day, and were instead deferred to return at a later date.

Deferral is an undesirable outcome known to discourage potential blood donors from returning to donate again. We would like to investigate the impact of the menopause on deferral (i.e. not being able to donate blood due to failing the test for blood haemoglobin), and whether this is influenced by iron supplementation. An analysis of this relationship should account for confounders of the association between deferral and menopause status and should include an investigation of whether this association differs by iron supplementation.

Variables in the dataset

Deferral is defined as a binary variable indicating whether a woman donated blood (coded 0) or was deferred due to failing the test for blood haemoglobin (coded 1). Menopause status (also binary) and a number of other factors (including iron supplementation) that may influence donation/deferral are also included in the dataset:

Name of variable	Description
id	Unique patient identifier
age	Age in years
bmi	BMI (kg/m ²)
smoke	Smoking status (0 = never smoker, 1 = ex / current smoker)
redmeatfreq	Red meat consumption (0 = never, 1 = 0-1 times/week, 2 = 2-3 times/week, 3 = 4-6 times/week, 4 = 7+ times per week)
fruit	Fruit consumption (0 = never/rarely, 1 = regular consumption)
tea	Tea consumption (0 = never/rarely, 1 = regular consumption)
menopause	Menopause indicator (0 = no, 1 = yes)
ironsupp	Iron supplementation indicator (0 = no, 1 = yes)
deferral	Deferral indicator (0 = donated blood, 1 = deferred)

Research questions

Your AA should answer two research questions:

Does the menopause increase the prevalence of blood donation deferral?

Is the effect of the menopause on deferral modified by iron supplementation?

Your aim is to develop a logistic regression model which estimates the effect of the primary exposure on the outcome. In your report, you should assess and discuss to what extent any of these variables act as confounders. Some factors may also potentially modify the effect of menopause on deferral. In this analysis, we are only interested in effect modification by iron supplementation, and you need only test and report this interaction.

Guidelines on how to write your report

Your report should be structured with sections on Introduction, Methods of Analysis, Results, and Discussion (including conclusions). The majority of the marks are allocated to the Results section which should cover descriptive analysis, model development and assessment of confounding and effect modification. You should aim to use around a quarter of your report describing your analysis strategy, around half on the results and about a quarter discussing your results.

The report should be understandable to people who have not been closely involved in the fieldwork but who are aware of the general aims of the study. Your report should use statistical and epidemiological methods that you have learnt in this module (EPM202) and demonstrate that you have appropriately applied these to the research question and can correctly interpret results. However, you do not need to explain the theoretical concepts behind the epidemiological or statistical methods that you use. The report should describe:

- 1) how you analysed your data, including the method that you used
- 2) the results you obtained
- 3) your interpretation of your results.

Many of the principles involved in writing a good paper for a scientific journal apply to this assignment but:

- 1) There should be **more detail on the strategy you used to analyse the data** than is found in most papers.
- 2) You should include a very brief introduction including a statement of the research question, why it is important, the study population, the outcome, the primary exposure and potential confounders. This introduction should be approximately 100 words.
- 3) You should not cite or discuss relevant literature but you should consider the internal validity of the study (and hence your results).
- 4) You should discuss the interpretation of your results and how they relate to the research question. This may include alternative statistical explanations for any associations you find.
- 5) You may assume that the techniques you have used (e.g. chi-squared test, Mantel-Haenszel methods, 95% confidence intervals, likelihood ratio test, etc.) are understood by the reader and do not need explanation, but you do need to say which methods you have used.

The following should be clear to the reader:

- 1) How you structured your analysis (what is the primary exposure, which variables were considered to be potential confounders or effect modifiers).

- 2) How you handled missing data.
- 3) How you grouped observations into categories (if you did), and whether there is a linear trend across exposure levels, or categories.
- 4) What analyses you performed (you may need to be selective in what you present) – your description of the analysis should be sufficiently clear that the reader could reproduce your results.

General guidelines and advice

Before completing this AA, it is advisable to:

- Study up to and including Session 9 plus Session 12 of the EPM202 CAL material.
- Work through Workbook Practicals 4 and 7-9 in Stata so that you are familiar with the appropriate commands to be used.
- Complete the EPM202 FA (formative assignment).

Other guidelines:

- Don't include variable names or unedited Stata output in your report.
- Don't include formulae or Greek letters.
- Don't present an unnecessary number of decimal places.
- Don't write "p=0.000".
- Start with simple analyses. Don't jump straight to regression analyses before first using simple methods to understand your data.
- Take the study design into account when reporting results.
- Explore all the exposure data given. You should explain your decisions on which variables to include and exclude from your analyses.
- State how missing values are dealt with and what implications this might have, if any, for the results you obtained.
- Show clearly how many observations (n) have been included in each of your analyses.
- When comparing crude and adjusted models, or comparing two models using a likelihood ratio test, make sure the models contain the same observations.
- Make sure tables have titles.
- Make sure tables can be understood without reading text.

- Try to ensure main argument of text can be understood without reference to tables and figures.
- Present stratum-specific results (if necessary).
- Try to avoid duplication between (and within) the three sections (methods, results, discussion) of your report.

You may use the Moodle EPM202 AA discussion forum (opens January 2020) to discuss conceptual issues, but this forum should not be used for sharing individual conclusions or answers (e.g. tables of results) and your submission must be your own work. **A report with any plagiarised work will be penalised.** Please see our Assessment Irregularities guidance (pages 2-32 of the [Academic Manual Chapter 7](#)) and other [plagiarism resources](#) for more details.

Criteria for grading

The criteria to be used in grading are as follows:

- 5** An outstanding report describing a very thorough analysis, clearly explained and appropriately interpreted, showing in-depth understanding of the analyses performed, and including a comprehensive discussion of the findings of the analysis.
- 4** A thorough analysis, well explained, with all major points addressed; but less thorough and/or less clearly explained than required for a grade 5, and/or with more limited discussion of the findings
- 3** A sound and generally thorough analysis, but some pertinent (relevant) points are omitted and/or minor errors made, and/or the presentation lacks clarity.
- 2** Understanding of most major points is shown, but some non-minor errors in the analysis or interpretation, or muddled presentation.
- 1 (fail)** Inadequate analysis and lack of understanding of major points/important concepts shown.
- 0 (fail)** Serious lack of understanding shown: inappropriate analysis used or serious misinterpretation of results.

Please note that a minimum grade of 1 is needed in the AA in order to gain any credit for this element of the module.

Submission format

Your report should be **no longer than 1500 words** and should be presented as double-spaced text, plus at most 3 tables, which should be presented on separate pages following the text. You should use the Arial 12-point font and margins should be at least 2 cm. The word counting facility in Microsoft Word should be used to check the number of words in your assessment. Please see the new [Assignment Penalties](#) document for details on the penalties for exceeding the word count but in summary:

- **Assessed assignments up to and including 10% over the specified word count will be deducted 1 grade point**
- **Assessed assignments of greater than 10% over the specified word count will not be marked and will be given an automatic zero grade.**

When submitting your report, **you should include your Stata log file**. Markers are not expected to routinely review your log file, and extra marks will not be given for results in the log file. However, the log file may be used to help inform the tutor feedback to you, if, for example the marker could not follow your analysis strategy or findings in the text, but could see the strategy in your log file. Please note though that the log file should be confined to the analyses discussed in your report.

When uploading your log file, you must zip this up with your AA as the AMS can only accept one attachment. Please also ensure that there is no identifiable information to you in the log file (or AA) as your work must be completely anonymised to the tutors e.g. delete the directory name or information that is generated at the start / end of creating the log file.

Submission deadline

The deadline for submission of this AA is **23.59 UK Time, 30th April 2020**. Please note the **new penalties for late submission** i.e. a reduction of one grade is imposed if the AA is submitted up to 48 hours after the deadline. **From 48 hours after the deadline, AAs cannot be submitted.** See the [Assignment Penalties](#) document for more details.