Faculty of Education, Health and Human Sciences

School of Human Sciences

Module title	Applied Epidemiology and Statistics
Module code	PSYC-1115
Module leader	Nadya Belenky
Level	7
Coursework Title	Report
Weighting (%)	100%
Submission details	Via Moodle through Turnitin under Coursework Submission
Submission deadline	8 January 2025
Return date	31 January 2025

1. Guidelines for the assessment

Each student will be provided with a dataset and will analyse the data, interpret the results and prepare a technical report of the findings using Stata. Each dataset will be slightly different, which means every students' results will be slightly different.

The Basics of the Analysis:

- Create descriptive statistics of the sample & descriptives of the main outcome variable
- Create inferential statistics to study the association between the main outcome variable and other variables
- Write an extended technical summary about your findings

<u>You (by yourself) will decide on the appropriate methods using the knowledge you've learned in this module. There are MANY ways to do this analysis correctly. But there are also MANY ways to do the analysis INCORRECTLY, and even more ways to interpret your output incorrectly.</u>

You must carefully consider which Stata output is necessary to include in the technical report. **Do NOT cut and paste tables directly from the Stata output.** You will need to create your own summary tables based on the Stata output. Your tables and charts should be reader friendly. **You ARE allowed to copy FIGURES (e.g. boxplot, histograms) directly out of Stata** because they are high-quality and professional-looking.

Rounding: Rounded values for your data to no more than four significant figures after the decimal point. Except for very small p-values, values of most statistics should be rounded to two decimal places.

Completing the assessment without plagiarism

While you may wish to discuss ideas with other students about how to analyse the data, **it is absolutely essential that you write up your results individually**. **Everything in the report must be your own, original work. No copy-pasting from other internet sources or academic papers, no copying other students' tables.** Working together on your written work or passing off other people's writing as your own is cheating and is considered academic misconduct.

You will most likely choose to analyse your data and present your findings in different ways from your classmates and there is no single correct approach. This also means that if two people have very similar layouts and methods, it is very statistically unlikely that the students haven't been working together. It is also very obvious to an instructor when two people have worked together. Do your own analytic work and your own writing.

If there is overlap between your code and another student's code OR between your writing and another student's writing OR between your writing and the internet: your assessment will be reported to the academic misconduct team and you run the very real risk of getting a zero on this assessment.

2. Analysis Instructions (yellow highlighting is specific for you)

Your outcome variable (and the focus of your technical report) is **Body Mass Index (BMI)**.

Descriptive statistics

Summarise the characteristics of the sample in a traditional "Table 1". Your Table 1 should have one column with ALL sample participants and then additional columns focusing on the sample characteristics across the categories of your outcome variable.

In your descriptive statistics table, you make the decision about how to display your variables. For example, if you have a numerical, continuous variable like age, you could produce descriptive for a continuous age variable OR you could create a new categorical age variable and create descriptive statistics for the new, categorical variables. These are YOUR choices to make.

Explain your choices using words. Explain your tables using words. **Your descriptive statistics have to be written out and explained** – you should be highlighting the characteristics of the sample that you think are most interesting.

Please create **descriptive statistics for the following variables for the FULL SAMPLE (everyone)**:

- Age
- Race
- Sex
- Education
- Health status
- Rurality
- Household size
- Prior heart attack
- Vitamin C levels

Next, please **create a categorical version of your outcome variable, BMI.**

Please use these cut-offs to create your new categorical BMI variable:

- "normal" if numerical bmi is less than 25
- "overweight if numerical bmi is greater than or equal to 25 AND less than 30
- "obese" if numerical bmi is greater than or equal to 30

Next, **create the same descriptive statistics for the same variables in the list above within each level of your categorical BMI variable (Normal, Overweight, Obese)**.

Finally, **make 2 figures.** Figure 1 should be a **histogram of your numerical outcome variable**. Figure 2 should be a **boxplot of your numerical outcome variable across a categorical variable of your choosing**.

Describe both figures in your own words.

Inferential statistics

In your dataset, you will be given a main outcome variable – which is the important health event that your report will focus on. For the technical report, I will ask you to use statistical tests to investigate several relationship, always between the main outcome variable and another variable.

Your outcome variable is **BMI**. Your inferential statistics may test for statistical significance with **numerical BMI OR categorical BMI**. Your choice.

Please **choose FOUR of the variables below** to **test for a relationship with BMI (numerical or categorical)**:

- Age (numerical or categorical)
- Rural
- Sex
- Race
- Education
- Vitamin C levels (numerical or categorical)

Ask yourself: which variables do you think have interesting relationships with BMI?

You will need to do FOUR hypothesis tests, one for each variable you select from the list above. **You choose what hypothesis tests to use based on what you've learned and how you've coded your variables.**

Each time you test a relationship between the outcome and another variable, you will need to:

1) specify what test you are using (AND WHY), on what variables, and how those variables are coded (categorical or numerical),
2) write out your null hypothesis,
3) show the results of the test (ideally in a table),
4) indicate whether you accept or reject the null hypothesis and what your decision was based on,
5) interpret your result.

3. Writing Your Report: What should my sections look like and what should go in them?

Technical reports are briefing documents written by technical experts on specific topics for decision-makers (often civil servants, programme managers or administrators.) They are commissioned to provide information on specific questions or issues and to provide a basis for decision- making and action. As such they should put forward all the relevant facts and set out the relevant issues. The aim is to inform the reader sufficiently to enable them to understand the reasons for and implications of any decisions and subsequent actions that they take.

Information presented in the report should not include personal views that are not supported by the data or by other evidence/literature. You should assume that **the person you are writing for is intelligent and proficient, but busy**, and not an expert in relation to the issue in hand and does not know how to code in Stata. The report should include a short background section, methods used, key results, limitations of your findings and policy recommendations.

Introduction (10% of your grade)

- Why the main outcome is an important issue
    - How common is the main outcome in the general population/world/country? Pinkland is an imaginary country, so you can base your background statistics on a real country, like the UK.
    - What other consequences does the health outcome have?
    - If you can find it: are there burdens on the health system because of this health outcome? What about economic burden?
- Why is this technical report needed? What are you going to explore?

Methods (15% of your grade)

- Brief description of data collection and sampling procedure
    - See **Assessment Data and Data Collection Description** document on Moodle
- Important features of the study design and quality control
    - See **Assessment Data and Data Collection Description** document on Moodle
- Description of how the variables used in the analysis were defined
    - This is where you talk about your decisions to recode variables (what you did, why you did it)

- Description of the analysis plan for descriptive and inferential statistics (for categorical and continuous variables)
    - What kinds of descriptive statistics will you produce?
    - What types of figures will you make? AND WHY?
    - What variables will you analyse?
    - What types of hypothesis tests will you use? AND WHY?
    - What statistical thresholds you will use to interpret your hypothesis tests

Note: you will need to use the document on data collection and description to understand the dataset, as well as to understand the original survey design, data sampling, and quality control.

Results (55% of your grade)

- A classic "Table 1", well-formatted, easy to correct, with the correct numbers in it
- Written description of the sample including descriptive statistics for the full sample as well as the descriptive statistics compared across the levels of the outcome variable
- Presentation of the inferential results: complete, unambiguous, correctly interpreted, all the key pieces for the correct statistical tests

Limitations (10% of your grade)

- Identify any limitations of your data analysis and the survey methodology (think about what you've learned about sampling, representativeness, and bias).

Policy Recommendations & Next Steps (10% of your grade)

- Identify the need for interventions to address the problems identified and **make recommendations that are specific to your results**
- You can be creative in this section: based on your results, what do you think should be the next steps? Policy-wise? Research-wise?

References

- Keep a list of all references in the Harvard format. Preferably use appropriate reference manager software for this.
- Your reference section **does not count towards your word count**.
- Points deducted from the sections that your references are in, if your references are 1) fake, 2) not relevant, 3) not academic, or 4) otherwise not good.

IMPORTANT IMPORTANT: Appendix with Stata Code

- Please copy and paste your .do file code at the end of your technical report
- This is good transparency practice and it also prevents plagiarism/cheating
- The Stata code **does not count towards your word count**
- Stata code appendix is required and you lose points if you don't include your code
- Points deducted from your results section if your code is not included or if your code is 1) fake, 2) non-functional, 3) clearly not related to your results.

4. **Length (max 2500 words +/- 10%)**

The maximum length of the report is not limited to the text only, this includes tables and figures/charts, but excludes the reference list and excludes the Stata code. **To simplify the length estimation, each table (regardless of the size) will count as 100 words and each figure as 50 words.**

For example, if you include 4 tables (4 x 100 = 400) and 2 graphs (2 x 50 =100) in your report you have used 500 words. The remaining 2000 will be distributed as text across the report.

Marking Criteria

The criteria for passing this assessment include:

- Appropriate analysis
- Appropriate presentation of the results
- Conclusions are based on the findings presented
- The style of the text is clear, simple, concise, logical and systematic

Data analysis and interpretation

- Use of appropriate tests
- Appropriate use and display of tables and graphs
- Presentation of results
- Adequate interpretation of key results
- Explanations clear and understandable
- Limitations of survey methodology and of your data

Academic writing and referencing

- Clarity and logical organisation of the report
- Degree of synthesis / creative thought demonstrated
- Page style / font / margins are appropriate
- Reference list and in text references consistent
- References using Harvard style
- Reference list complete and without errors
- Supplementary items cross referenced and appropriate
- Appropriate text explaining tables and graphs
- Clear English with coherent flow and correct grammar
- Appropriate length