

ASSIGNMENT 2A: ASSESSING DATA QUALITY AND PRODUCING REPRODUCIBLE NOTES

Group 2

Hi Angela, Katherine, Hossain and Zhenyu,

This assignment assesses students' knowledge and skills in evaluating data quality, documenting findings and decisions, producing reproducible notes as well as understanding of data linkage strategies and best-practice privacy preserving procedures. This assessment provides opportunities for students to consolidate acquired knowledge about data management and analysis and further strengthen critical reflection.

Below are overall grade (highlighted in green) and detailed scores for the work group. In a separate attachment, you can find detail feedback and comments provided alongside your answers and/or in track changes. Final mark for everyone (considering Peer Assessment) can be found in the Open Learning "Your progress" section as usual.

Best regards,

Dr Nicole Mealing

HDAT9400 Lecturer, T2 2023

Marking rubric for quality of Assignment 2A

The grading rubric is outlined below.

Criteria	HD 85 – 100	D 75 – 84	Credit 65 – 74	Pass 50 – 64	Fail Less than 50
Written documentation of reproducible notes – data dictionary, cleaning notes, flowcharts	Full logic shown with elaborate and very coherent descriptions which goes beyond replication	Full logic shown with coherent description	Most logic shown with mainly adequate detail	Some documentation presented but logic not clear	Not presented
Question answers	All correct answers, full working shown, with in-depth reflection and suggestions	All correct answers, full working shown, with some critical reflection	Mostly correct answers, with some misinterpretation, working shown	Incomplete answers, misinterpreted results in most parts.	No answer
SAS code	Fully functioning and reproducible code with extensive and consistent annotation and commenting throughout	Fully functioning and reproducible code with moderate but consistent annotation	Functioning code with some annotation	Code with some errors, lacking annotation.	Code with errors in most parts, or no code provided

The scoring rubric for quality of Assignment 2A is outlined below.

Assessment of written document* and SAS code**	Max score	Your mark
Part 1. Documentation of data cleaning decision		
1.A – GP dataset exploration and cleaning of (25 marks)		
A written document* explaining results of exploration, decisions made, and results of GP data cleaning	10	7
Flow chart for GP data cleaning	5	4.75
SAS code showing work for GP data exploration, decisions made and producing results of cleaning	10	9
1.B – ED dataset exploration and cleaning (20 marks)		
A written document* explaining results of exploration, decisions made, and results of ED data cleaning	8	4.5
Flow chart for ED data cleaning	5	4.5
SAS code showing work for ED data exploration, decisions made and producing results of cleaning	7	5.5
1.C – SAS code to create 4 new variables (10 marks)		
Smoke_status_GP (including documented justification)	6	5
Risky_alcohol_GP	1	0.8
BMI_GP	1	0.8
Obese_GP	1	0.8
HighBP_GP	1	0.8
1.D – Update data dictionaries (10 marks)		
Updated data dictionary for GP dataset	5	5
Updated data dictionary for ED dataset	5	4
Part 2. Analyse cleaned GP data and report (20 marks)		
Tabular presentation of results	7	5.8
Written interpretation of results*	8	7
SAS code to generate results	5	4.2
Part 3. Data linkage (15 marks)		
Data linkage strategies	5	5
Diagram depicting variable exchange information	10	8.55
TOTAL MARK	100	83

*: Assessment of written document includes the use of academic and linguistic conventions, coherent and succinct discussion with supporting information presented.

**:. Assessment of SAS code include i) correct use of common SAS commands introduced in the course, ii) effective creation of variables that takes into account variable type and values, and iii) clear annotation that enables reproducibility.