University of Glasgow | College of Medical, Veterinary & Life Sciences

# Master of Public Health

## Further Epidemiology and Statistics (MED5021)

## *Statistics Re-assessment 2022-23*

**Attempt all questions. Word count 2,000 words. Please note, there are penalties for exceeding the word count for assessments. Information about these are available in the assessment section on the course Moodle site.**

**This coursework is worth 50% of your overall mark for this course.**

**Submission deadline – 2nd August 2023**

**TOTAL AVAILABLE MARKS= 50**

It is expected that you will analyse the data using **R or STATA**.
The data sets are available on the course Moodle site.

When analysing the data and writing your report, you should:

- provide ***appropriate*** plots and/or tables to summarise the data,
- state and justify any assumptions you make, and state any reservations you may have about these assumptions,
- describe the analysis you have done
- interpret the results robustly and in non-technical language

Your report should include the ***relevant*** sections of R/STATA output, copied and pasted into your Word document. You should ensure that your output and variables are appropriately and informatively labelled. [Do **NOT** print out and hand in the entire contents of your R/STATA output viewer – only include ***relevant*** output.]

*Please do **not** analyse the data from every conceivable angle! A well-considered and justified analysis with clear interpretation is necessary.*

***(continued over page)***

**Question 1**

A sample of first year male statistics students at Glasgow University had the grip strength of their dominant hand measured using a grip dynamometer. In addition a number of physiological measurements were made on their dominant hand and arm; width and length (cm) of the hand, circumference (cm) of forearm and bicep, skinfold thickness (mm) of forearm and bicep. The height (cm) and weight (kg) of the students was also recorded.

The data are stored in the Stata data file **Male grip strength.dta** and contains the following variables:

| | |
|---|---|
| **Hwid** | hand width (cm) |
| **Hlen** | hand length (cm) |
| **Fcirc** | forearm circumference (cm) |
| **Bcirc** | bicep circumference (cm) |
| **Fskin** | forearm skinfold (mm) |
| **Bskin** | bicep skinfold (mm) |
| **Grip** | grip strength (kg) |
| **Weight** | weight (kg) |
| **Height** | height (cm) |

(a) Use the weight and height variables to create 4 categories of BMI. (i) underweight <18.499 (ii) Normal 18.5-24.999 (iii) overweight 25-29.999 and (iv) obese >=30 and then use appropriate statistical methods to identify a parsimonious model for predicting grip strength using some combination of the explanatory variables.

***(10 marks)***

(b) For your chosen model, provide a clear interpretation of the coefficients in the model and indicate how the model might be used to predict the grip strength of a future male student.

***(10 marks)***

(c) Check the assumptions underlying your model.

***(5 marks)***

***(continued over page)***

**Question 2**

You have been asked by an endocrinologist to develop a statistical tool that predicts the probability of a patient presenting with type A or type B diabetes from patient clinical and lifestyle data. She would like the tool to be as parsimonious as possible and be able to apply it to males and females separately.

The data are stored in the Stata data file **diabetesmodel12_corrected.dta** as follows:

| | |
|---|---|
| **id** | patient identifier |
| **age** | patient age in years |
| **bmi** | patient bmi in kg/m2 |
| **chd** | patient has coronary heart disease Yes/No |
| **chol** | patient cholesterol level mg/100 ml |
| **dbp** | diastolic blood pressure |
| **diabetes** | patient diabetes status: type A or type B |
| **ncigs** | number of cigarettes smoked per day |
| **sbp** | systolic blood pressure |
| **smoke** | patient smokes: Yes/No |
| **Sex** | sex of patient: 1= female, 2= male |

a) Use appropriate tables, plots and summary statistics to explore the effect of age, bmi, and sbp on diabetes for the *whole dataset*.

***(7 marks)***

b) For males and females *separately,* build a parsimonious logistic regression model that predicts patient diabetes type from the available variables and interpret all coefficients in the final models.

***(14 marks)***

c) Check that each model fits the data.

***(4 marks)***

***NOTE: PLEASE DO *NOT* COPY AND PASTE NUMERICAL OUTPUT, OR SYNTAX, DIRECTLY INTO YOUR COURSEWORK. IF YOU DO, THEN YOU MAY LOSE SOME MARKS BECAUSE OF POOR PRESENTATION.***

**[END OF PAPER]**