Analysis of Cardiometabolic Risk Factors in Multicultural Communities (total: 100 marks)

Refer to the **Assessment Criteria and Guidelines** in the Assessment block of the Mylearn site. **Five marks** will be awarded for the following:

- Clear, concise expression including correct use of terminology and notation.
- Presentation of figures and tables including relevant R output.
- Clearly and concisely annotated R code (in a script file or Rmd file).

NOTE: For **all** analyses you should

- explain your model selection process, including relevant output.
- adjust for overdispersion, if it exists.

Background and dataset

Some migrant groups have higher risks of deaths and chronic diseases due to barriers associated with socioeconomic disadvantage, social isolation, racism, language, poor access to health services, and low levels of health literacy. However, few culturally-tailored interventions have targeted ethnic groups in Australia.

In this assignment, you will be analysing data from the Living Well Multicultural – Lifestyle Modification Program, a health intervention study conducted in Queensland, Australia between October 2014 and June 2017. This program was designed to address cardiometabolic health disparities in ethnic communities through tailored lifestyle interventions. These articles below will be helpful for your understanding of the dataset; however, the dataset used in the assignment has been modified, so you will not obtain the same results as reported in the papers.

1. Gallegos, D., Do, H., To, Q.G, Vo, B., Goris, J., Alraman, H. (2021). The effectiveness of living well multicultural-lifestyle management program among ethnic populations in Queensland, Australia. *Health Promotion Journal of Australia*, 32 (1).

2. D Gallegos, H Do, QG To, B Vo, J Goris, H Alraman (2019). Differences in cardiometabolic risk markers among ethnic groups in Queensland, Australia. *Health & social care in the community*, 27(4).

The dataset contains baseline measurements from adults ($\geq$ 18 years) from ethnic communities in Queensland who were not underweight and eligible for the lifestyle program.

Response variables

- Body Mass Index (BMI) - continuous
- Waist-to-height ratio (WHtR) - continuous

- Hypertension status - binary (for hypertension analysis)

Independent/Predictors variables (Table 1). After recoding the variables as instructed in Table 1, you will need to create 2 new variables.

- **Eating**: Sum all eating behaviour variables (fruit, veggie, milk, takeaway, Chips, salty snacks, sweet snacks, Softdrink, Pr_meat), into a single score, ranging from 0 to 9.
- **Physical**: Sum all the physical activities (Walking, moderate activity and vigorous activity) into a total activity time with vigorous acitivity doubled. If the total ≥ 150 minutes, then code **Physical** as "meeting guideline", otherwise it is "not meeting guideline".

Important: Use only Eating and Physical in your models, not the original raw variables.

Statistical Analysis and modelling

Question 1: Exploratory analysis **[15 marks]**
Conduct an exploratory analysis to familiarise yourself with the data before modeling. Examine the distribution of the response variables, explore relationships with potential predictors, and identify any data quality issues. Provide visualisations and summary statistics to support your findings.

Question 2: Cardiometabolic risk modelling **[50 marks]**
You model cardiometabolic risk using both BMI and WHtR as response variables in separate analyses. For each outcome:

- Perform model selection and/or variable selection using appropriate methods (e.g., stepwise regression, information criteria like AIC/BIC, or hypothesis-driven approach)
- Compare competing models and justify your final model selection.
- Present results for your selected model with interpretation of coefficients and confidence intervals.
- Check model assumptions and perform diagnostics.
- Compare the findings between the two models and discuss which anthropometric measure appears to be more strongly associated with the predictor variables
- Provide a summary comparing the two analyses and their implications for understanding cardiometabolic risk.

Table 1: Description of variables in the *LivingWell.csv* dataset

Variable	Description
Community	categorized into five main groups based on geographic locations including: Burmese/Vietnamese (Southeast Asia), Sri Lankan/Bhutanese (South Asia), Afghani/Arabs (Middle East), Somalian/Sudanese (Africa), and Pacific Islander.
Dose	the number of sessions that the participant attended (from 1 to 8). You should recode Dose to "completed" if it is ≥ 7 sessions, otherwise it is "incomplete".
Age	age of the participant at the time the program started (years)
Gender	categorised as "female" or "male".
Time_aus	Time in Australia was self-reported by participants. You should recode it to "long-stay" if it is more than 5 years, otherwise it is "short-stay".
Edu	education categorised into 4 groups "primary school", "high school", "Trade/certificate", or "bachelor/postgraduate degree".
Employ	Types of employment were "paid work", "work without pay", "retired/unable to work", or "student".
House	Household types
Smoke_stt	smoking status (Yes/No)
Fruit	Fruit consumption per day. You should recode it to 1 if the participant consumed ≥ 2 pieces per day, and 0 otherwise.
Veggie	Veggetable consumption per day. You should recode it to 1 if the participant consumed ≥ 5 servings per day, and 0 otherwise.
Milk	Type of milk the participant consumed. You should recode it to 1 if it is "low fat", and 0 otherwise.
Takeaway	You should recode this variable to 1 if the participant consumed $\leq$1 time per week, and 0 otherwise. Apply the same recoding to: **Chips, salty snacks, sweet snacks, Softdrink, Pr_meat**.
Walking	time spent on walking (minutes)
Md_act	moderate activity (minutes)
Vigorous_act	vigorous activity (minutes)

Question 3: Hypertension modelling **[30 marks]**

Model hypertension status:

- Perform model selection and/or variable selection using appropriate methods (e.g., stepwise regression, information criteria like AIC/BIC, or hypothesis-driven approach)
- Compare competing models and justify your final model selection.
- Present results for your selected model with interpretation of of odds ratios and confidence intervals.
- Check model assumptions and perform diagnostics.
- Discuss the clinical/practical significance of your findings for hypertension risk in multicultural communities