**The University of Western Australia**

**School of Population and Global Health**

**BIOSTATISTICS II (PUBH5769)**

**Assignment 1 (Topics 1-4)** Due 4pm Monday 31 August 2020

Instructions

You must do the Assignment entirely on your own. You must not discuss the questions or answers with others.

Assignment should be typed. Note that you can copy & paste selected SAS syntax and SAS output into a Word document. For questions requiring use of SAS you must provide a copy of your SAS program (i.e. syntax) as well as the relevant output (and any requested additional commentary and hand-calculation). Do not hand in duplicated or unrequested output. Show working and reasoning. Write proper sentences.

All students must prepare a PDF copy of their Assignment and name the file AssignOne\_Yourstudentnumber.pdf where Yourstudentnumber = your student number.   
Upload the file to LMS Assignments by the due date.

Marking

Total marks for each question are shown. Marks will be deducted for incorrect and incomplete answers, inadequate explanation, poor quality comment and interpretation sentences, not following instructions and poor presentation.

Community survey dataset

This dataset (survey.sas7bdat) relates to a cross-sectional community health survey and contains data on 1552 participants aged 40-69 years. Use the community survey dataset for both questions in this assignment.

**Question 1** [10 marks] Use the community survey dataset to do the following.

1. [2 marks] Use Proc GLM to fit a linear model that compares mean DBP for adults with RXHYPER=yes vs no. Obtain the estimated difference in mean DBP for adults with RXHYPER=yes vs no, its 95% confidence interval and the associated p-value. Write a sentence that includes and interprets these results.
2. [2 marks] Using Proc GLMSELECT perform a stepwise (backward) search for predictors of DBP from the following list of potential predictors: age, sex, bmi, drinking, smoking, diabetes, and exercise. In your analysis consider sex, drinking, smoking, and diabetes as categorical (i.e. class) variables and the remainder as quantitative variables. In your search for predictors consider main effects and squares of quantitative variables only (i.e. not interactions) and use p=0.05 criterion for dropping variables. Provide the output that shows the order in which terms were dropped and the output showing the fitted final model (i.e. its estimated coefficients).
3. [3 marks] Obtain or calculate and interpret the estimated effect (on DBP) of each variable in your final model from (b). Which of the continuous variables in your final model in (d) have the largest impact on mean DBP.
4. [2 marks] Use Proc GLM to fit a linear model that compares mean DBP for adults with RXHYPER=yes vs no after adjusting for all the variables in your final model in (b). Obtain the estimated (adjusted) difference in mean DBP for adults with RXHYPER=yes vs no and its 95% confidence interval and the associated p-value. Write a sentence that includes and interprets these results.
5. [1 mark] Write a sentence that describes the impact of the adjustment variables on the comparison of mean DBP for adults with RXHYPER=yes vs no.

**Question 2 [10 mark]**

(a) [2 marks] Use Proc GLMSELECT to perform a stepwise (backward) search for predictors of FVC from among the following list of potential predictors: sex, age, bmi, smoking, height, exercise and asthma. In your analysis consider sex, smoking, and asthma as categorical (i.e. class) variables and the remainder as quantitative variables. In your search for predictors consider main effects and interactions with asthma (but do not consider squares of quantitative variables) and use p=0.01 criterion for dropping variables. Provide the output that shows the order in which terms were dropped and the output showing the fitted final model (i.e. its estimated coefficients).

(b) [4 marks] Use Proc GLM to fit the final model from GLMSELECT and obtain and interpret the estimated effect (on FVC) of each variable in the final model.

(c) [1 mark] Write down the algebraic representation of the fitted model from (b) for those with and without asthma.

(d) [1 mark] If your final model (from (b)) was used to predict FVC, what would be the estimated difference in predicted FVC for the following man and woman:

sex=male, asthma=no, smoking=never, age=50, height=1.8, bmi=25, exercise=7.  
sex=female, asthma=yes, smoking=never, age=50, height=1.7, bmi=25, exercise=7.

(e) [2 marks] Use hand calculations to obtain the approximate 95% prediction interval for FVC for the first individual in (d). If you used SAS to obtain these intervals would you expect them to be similar to your hand calculation, if so explain why.