# HSH746/HSH946 BIOSTATISTICS 1

# ASSIGNMENT 1 (20% OF TOTAL MARK)

## Due date: Friday 9 December 2022

### Instructions

Please note: this assessment task must be all your own work. Please do not discuss questions and answers in detail with your fellow students.

Assignments must be submitted on-line via the assignment folder in the unit site **by 8 pm on 9 December 2022**. Assignments must be submitted in a Microsoft Word document or an *editable* pdf.

Some of the questions may require calculations. The formula you use and your calculations should be included with your answers. If the final answer is incorrect, assessors can determine whether this is because of a simple calculation error (small loss of marks) or because of an incorrect formula or incorrect figures.

Some of the questions may require calculations using Stata. Where you have used Stata for calculations, you should copy the Stata commands and output from the Stata results screen and paste them into your assignment so that the assessor can see how you have derived your answer. Note: this Stata output is required in addition to your answer to the question. Simply pasting in the Stata output will not be considered an adequate answer on its own.

This assignment is worth 20% of the final mark for HSH746/HSH946 and the marks allocated for each question are shown.

Students should ensure that they keep a spare copy of their work.

Student Name:

Student ID Number:

# Questions

A Study in Norway examined the association between psychoactive substance use and sensation seeking behaviour among Norwegian drivers in Norway. An extract of the study sampling method is detailed below. The full journal article can be found in A1 folder on Cloud Deakin.

[The Norwegian study](https://link.springer.com/article/10.1186/s12889-019-8087-0)

A screenshot of a computer

Description automatically generated with medium confidence

Use the study above to answer **questions 1-2**.

**Questions**

1. Which of the following is correct about the target population in the above study

The target population is (choose 1)

Norwegian drivers in the arctic, rural county of Finnmark

Norwegian drivers from south-eastern Norway

Norwegian drivers

Norwegian drivers in the six counties of Finnmark, Oslo, Akershus, Buskerud, Hedmark and Oppland **(2 mark)**

1. Is the study representative of the target population? Give two reasons for your chosen response. **(2 marks)**

1. **Read the following data description and answer the following questions:**  
   The data set *assignment 1 heart data.csv* contains the results of a population study of heart disease and its associated risk factors among Australian adults. A complete description of the variables on the data set is contained in the word document *assignment 1 heart data description*. You should read this data description before attempting to answer the questions.   
     
   These are synthetic data, but you may reference them in your answers as coming from  
   *assignment 1 data: heart study.*  
     
   The data were collected on paper forms and transferred to a computer file using manual data entry.  
   1. What is your main task in preparing the data set for analysis? **(1 mark)**
   2. Perform the task you identified in part 1.

Show what you did – including any relevant Stata output. **(3 marks)**

*Note that all tables and graphs in this assignment should be presented with appropriate headings and footnotes*.

1. **Read the following data description and answer the following questions:**  
   A study collected heights in centimetres of 400 medical students. The data for this study is in the data set   
   *assignment 1 heights data.csv*  
   The variables on the data set are:  
   *index*: study identification number  
   *heights*: the students’ measured height in centimetres  
   You may cite the source of these data as  
   *assignment 1: heights data*  
   1. Plot a histogram of the heights with *frequency* on the y axis **(2 marks)**
   2. Using this histogram state which of the following best describes the heights variable. **(1 mark)** 
      1. The distribution of heights is left skewed
      2. The distribution of heights is symmetric; or
      3. The distribution of heights is right skewed.
   3. Calculated the appropriate measure of location and spread to 1 decimal place  
      (this will be either a mean and standard deviation or a median and interquartile range)   
      **(3 marks)**
   4. Assuming that this is a sample which is representative of all medical students, what is the height which we would expect 10% of medical students to be shorter than? State your answer to 1 decimal place. **(1 mark)**

1. **Read the following study description and answer the following questions:**

Suppose you wish to select a sample of Deakin university students to participate in a survey of their leisure time activities. You wish to use this sample to make inferences about leisure time activities for the general Deakin student population.

You go to an on-campus coffee shop and as students arrive during the day you record their names. At the end of the day, you choose the names of 20 students from this list by placing the names in a hat and drawing out 20 of them.

* 1. This sample is *(choose 1)*

1. A simple random sample;
2. A stratified sample
3. A convenience sample
4. A systematic sample. **(1 mark)**  
   1. What is the sampling frame? **(1 mark)**
   2. What is the target population? **(1 mark)**
   3. Identify two potential sources of bias coming from the way we have selected the sample.   
      **(2 marks)**

**End of assignment questions**