# HSH746/HSH946 BIOSTATISTICS 1

# ASSIGNMENT 1\_1 (20% OF TOTAL MARK)

## Due date: 11 December 2020

### 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 in Deakin sync **before 8 pm on 11 December 2020**. 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. ***Note that all tables and graphs in this assignment should be presented with appropriate headings and footnotes*.**

***Please submit two copies of the assignment to the assignment submission folder. The first copy should include the original assessment questions (to make sure you have answered all questions) and the second copy should exclude the original questions to check assignment originality with Turnitin.***

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:

“The monthly Labour Force Survey (LFS) provides information about the labour market activity of Australia's resident civilian population aged 15 years and over. The LFS is designed to primarily provide estimates of employment and unemployment for the whole of Australia and, secondarily, for each state and territory. The ABS has been conducting the Labour Force Survey since 1960, initially as a quarterly survey. In February 1978, the frequency of the survey was changed from quarterly to monthly”. The link to the study methodology is given below.

[Labour Force Australia](https://www.abs.gov.au/methodologies/labour-force-australia-methodology/oct-2020)

Use the ***ABS LFS link*** above to answer questions 1 & 2.

# Questions

1. Which of the following is most correct about the ABS LFS? **(1 mark)**

The LFS sample *(choose 1)*

1. Provides information on labour market activity of Australian residents
2. Provides information on labour market activity for each state and territory in Australia
3. Provides information on labour market activity of all Australian residents aged 15 years and over
4. Provides information on labour market activity of all Australian civilian residents aged 15 years and over
5. A and B
6. None of the above
7. What type of sampling method can be used to select the sample for LFS? Give 2 reasons for your chosen response **(2 marks).**
8. UNICEF data on Care seeking for diarrhoea indicates the percentage of children (under age five) with diarrhoea for whom advice or treatment was sought from a health facility or provider. ***A1\_DiarCare.csv*** is a sub-sample of the UNICEF diarrhoeal care data. The data contains the following variables:

|  |  |
| --- | --- |
| Year | Data collection year |
| Data source | Source of data used |
| Country | Average percentage at national level |
| Sex | Male |
|  | Female |
| Area | Urban |
|  | Rural |
| Wealth Index Quintile | Poorest |
|  | Second |
|  | Middle |
|  | Fourth |
|  | Richest |

# Use the *A1\_DiarCare.csv* data for question 3.

3.1. Considering all countries, what is the average percentage of children under five that sought advice or treatment from a facility or provider? Indicate the associated SD and IQR (p75-p25) and report to 2 decimal places (**1 mark).**

3.2. What is the nature of the distribution of the national variable data? Give stats to support your answer (**1 mark).**

3.3. Which of the following is most correct? (**1 mark)**

A: Children in urban areas always sought advice for diarrhoea care from a health facility or provider more than children in rural areas.

B: Children in urban areas sought advice for diarrhoea care from a health facility or provider more than children in rural areas by 46.06%

C: Children in rural areas sought advice for diarrhoea care from a health facility or provider less than children in urban areas by 44.92%

D: Children in rural areas sought advice for diarrhoea care from a health facility or provider less than children in urban areas by 4%.

3.4 Using the appropriate measure of spread and location, compare the poorest and richest quantiles for all countries **(3 marks).**

3.5 What conclusion can you make based on the results above for poorest and richest quintile seeking diarrhoea care for children? **(1 mark)**

4. The data set ***T3\_A1\_2020\_NHIS\_adult\_NCD.csv*** is a sub-sample of the 2019 NHIS adult sample (13 variables 6,997 observations). A complete description of the data variables is found in the word document ***T3\_2020\_A1\_nhis\_adultsampledata\_description***. You should read this data description before attempting to answer the questions.

Use ***T3\_A1\_2020\_NHIS\_adult\_NCD.csv*** data to answer question 4.

**For the following questions, use data from quarters 1 and 2 (variable invt\_QRT)**

For this question, you will use the variable ***phstat\_a***, which is the response to the question *"Would you say your health in general is excellent, very good, good, fair, or poor?"*

Explore the variable ***phstat\_a*** using graphical analysis - the graph should indicate percentage of frequencies on the Y axis.

4.1 Before drawing the graph - rename the variable ***phstat\_at*** as *health\_status* (**½ mark).**

4.2 Add value labels to renamed variable *health\_status* **(1 mark).**

4.3 Draw the graph for health status including graph bar heights **(2 Marks).**

4.4 Explain the results from the graph **(1mark)**.

1. The data set ***T3\_A1\_2020\_NHIS\_adult\_NCD\_datacleaningdata.csv*** is a sub-sample of the 2019 NHIS adult sample (9 variables 1,000 observations). See data description in the word document ***T3\_2020\_A1\_nhis\_adultsampledata\_description***.

Use ***T3\_A1\_2020\_NHIS\_adult\_NCD\_datacleaningdata.csv*** data to answer question 5.

5.1 This question is about data cleaning: Use data from **all quarters** for question 5 **(3 marks)**.

Check the variables for any errors using the tabulate command and correct the error appropriately.

Do consider variable association where applicable. Comment on the responses given and justify why data cleaning is necessary.

5.2. What is the proportion and number of respondents advised to take aspirin that are now still following the doctor's advice? **(1/2 mark).**

6.1 Generate a new variable ***pain*** as follows and add value labels **(1 Mark).**

|  |  |  |  |
| --- | --- | --- | --- |
| pain = 0 if paifrq3m\_a is reported as never (add value label ***none***) |  |  |  |
| pain = 1 if paifrq3m\_a is reported as some days (add value label ***mild***) |  |  |  |
| pain = 2 if paifrq3m\_a is reported as most days or everyday (add value label ***moderate/severe***) | | | |

6.2 Did pain impact men or women more? Use stats to support your answer and report to 2 decimals **(1 Mark).**