1. Enter data in Table 1 into SPSS.

(a) Provide appropriate names and labels for the variables and present a screenshot of the Variable view. (2 marks)

(b) Create a fifth variable named HDI\_level that groups the countries into two groups of High HDI (0.851 to 0.957) and Low HDI (0.459 to 0.674) and obtain a frequency of the new variable. (2 marks)

(c) What is the average mean years of schooling in the ten countries with high HDI? (2 marks)

(d) What is the average life expectancy for the ten countries with low HDI? (2 marks)

(e) Comment on the standard deviations for the average life expectancy in the two groups of countries. (2 marks)

**Table 1 HDI, Life Expectancy and Mean Years of Schooling in a selection of countries.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **HDI** | **Life Expectancy** | **Mean years of schooling** |
| Norway | 0.957 | 82.4 | 12.9 |
| Hong Kong | 0.949 | 84.9 | 12.3 |
| Denmark | 0.940 | 80.9 | 12.6 |
| New Zealand | 0.931 | 82.3 | 12.8 |
| Japan | 0.919 | 84.6 | 12.9 |
| Spain | 0.904 | 83.6 | 10.3 |
| Italy | 0.892 | 83.5 | 10.4 |
| UAE | 0.890 | 78.0 | 12.1 |
| Portugal | 0.864 | 82.1 | 9.3 |
| Croatia | 0.851 | 78.5 | 11.4 |
| Iraq | 0.674 | 70.6 | 7.0 |
| Namibia | 0.646 | 63.7 | 7.0 |
| India | 0.645 | 69.7 | 6.5 |
| Ghana | 0.611 | 64.1 | 7.3 |
| Kenya | 0.601 | 66.7 | 6.6 |
| Zambia | 0.584 | 63.9 | 7.2 |
| Uganda | 0.544 | 63.4 | 6.2 |
| Nigeria | 0.539 | 54.7 | 6.7 |
| Malawi | 0.483 | 64.3 | 4.7 |
| Yemen | 0.459 | 66.1 | 3.2 |

**Section B**

This section requires you to use SPSS to enter, manipulate and analyse data. Please present your work in word i.e. copy the relevant SPSS output and

You should include a syntax of all your SPSS commands as an appendix to your assessment submission.

To produce a syntax follow these steps:

Click of **File New Syntax** as shown in the picture below

The data are available on Moodle in the Artefact 1 Assessment section. The data file contains information on 8,165 women aged between 15 and 49 years at the time of the survey. There are 11 variables, two of which – **BMI** and **Age** --‐ are continuous, and the other nine – **Age\_group**, **Province, Residence, Education, Water\_source, Wealth\_Index, Breastfeeding\_status, Anaemia, Marital\_status and Occupation**, --‐ are categorical.

A full description of the variables, including the coding scheme used, is shown in Table 2 below. (You can also click the "Variable View" tab from the Data Editor window in SPSS and then look at the "Label" and "Values" columns.).

Table 2. Variables and Coding Scheme.

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| **BMI** | Body Mass Index  A Body Mass Index is a measure of health/nutrition status which is calculated by dividing weight by height squared. So for example an individual with a weight of 60kg and a height of 1.7metres would have a BMI of 21 (60/1.72). This measure is applied to adults only. |
| **Age** | Woman’s age in years (ages from 15 to 49 years) |
| **Age\_group** | The Age group variable is coded:  1 = 15 - 19  2 = 20 - 24  3 = 25 – 29  4 = 30 – 34  5 = 35 – 39  6 = 40 – 44  7 = 45 - 49 |
| **Province** | The Province variable is coded:  1 = Kinshasa  2 = Bandundu  3 = Bas-congo  4 = Equateur  5 = Kasai-occidental  6 = Kasai-oriental  7 = Katanga  8 = Maniema  9 = Nord-Kivu  10 = Orientale  11 = Sud-Kivu |

|  |  |
| --- | --- |
| **Residence** | The Residence variable is coded:  1 = Urban  2 = Rural |
| **Education** | The Education variable is coded:  0= No education  1= Primary education  2 = Secondary education  3 = Higher education |
| **Wealth\_index** | The Wealth\_index variable is coded:  1 = Poorest  2 = Poorer  3 = Middle  4 = Richer  5 = Richest |
| **Breastfeeding\_status** | The Breastfeeding status variable is coded:  0 = No  1= Yes |
| **Anaemia** | The Anaemia variable is coded:  1= Severe  2 = Mild  3 = Moderate  4 = Not anaemic |
| **Marital\_status** | The Marital status variable is coded:  0 = Never married  1 = Married  2 = Living with partner  3 = Widowed  4 = Divorced  5 = No longer living together/separated |
| **Occupation** | The Occupation variable is coded:  1 = Not working  2 = Non-agricultural work  3 = Agricultural work |

1. Recode the variable ‘Age group' into a new variable called 'Age\_group2' with three categories ( 15 to 24 years, 25 to 39 years and 40 to 49 years). Make sure that the recoded variable is clearly labelled. Obtain frequencies of the new variable Age\_group2. What proportion of women are aged 25 to 39 years and what is the percentage of women in the age group 40 to 49 years? (4 Marks).
2. Recode the Variable Anaemia into Anaemic with two categories; Anaemic and Not Anaemic and produce a frequency of the recoded variable. (2 marks)
3. Recode the Variable Marital\_status to Marital\_status2 with two categories; one category should represent those that are married or are living with a partner and the other category should represent the unmarried or those without a partner and produce a frequency of the recoded variable. (2 marks)
4. Obtain and present the mean BMI and the 99% confidence interval for the mean BMI for each category of the Occupation variable **to two decimal places**. Based on the confidence intervals does the mean BMI differ significantly by Occupation? Interpret the confidence interval for the BMI of women that are not working. (6 marks).
5. Conduct a statistical test to explore if there are differences in the mean BMI for women that are Anaemic and those that are not Anaemic? What is the appropriate test to be carried out? Test at the 5% significance level. Make sure you state your hypotheses clearly, give the value of the test statistic and present your conclusions based on the result of the test. (7 marks).
6. Is there an association between a woman’s residential status (Residence) and their education level (Education)? What is the appropriate test to be carried out? Carry out the appropriate test at the 1% significance level to answer this question, stating your null and alternative hypotheses clearly. If an association exists, what is the nature of this association? (6 Marks)
7. Undertake a multiple linear regression analysis in SPSS using the **enter** method to identify the best model of the factors associated with BMI for women in the Democratic Republic of Congo. Your response variable is BMI (Body Mass Index) and the potential explanatory variables are all other variables available in the data set**.** 
   1. Identify a selection of explanatory variables for inclusion in the linear regression model. Your identification of explanatory variables should make reference to literature sources on factors/correlates of under/over-nutrition/obesity of which BMI is a measure of. *Your answer should be presented in a paragraph and you should include an in-text citation of sources that support your arguments following APA 7th Referencing style*. *A References list should also be provided.* **(Maximum 500 words**). (10 marks)
   2. Compute and present the frequencies of all dummy variables that are potential explanatory variables for the multiple linear regression model.( 6 marks)

c. Run a multiple linear regression, present and Interpret the SPSS output from the multiple linear regression analysis.(5)

1. Undertake a residual analysis to assess if your final model fulfills the assumptions of a linear regression model. (5 marks)
2. Discuss the results of the multiple linear regression analysis, stating whether there is consistency or not with previous studies. Where there is inconsistencies explain the potential sources of the inconsistencies. **(Maximum 500 words**). You need to cite references and provide a References list using the APA 7th referencing style, see: (8 Marks).