

The University of Western Australia
School of Population and Global Health
PUBH4401 - BIOSTATISTICS I

Assignment 1 (Topics 1 and 2) Due 5pm Tuesday 18 August 2020 **[Total: 20 marks]**

You must do the assignment on your own. Do not discuss the questions or answers with any other person.

The assignment should be typed - you can copy and paste selected SPSS output into a Word document. Show all working and reasoning. Do not hand in duplicated or unrequested SPSS output. Marks will be deducted for inadequate explanation and poor presentation.

Students must submit completed assignments as one document **in PDF format** using the submission system on the unit LMS pages by 5pm on the due date. For this assignment, this will involve uploading only one completed PDF document to LMS. You must name the file AssignOne_Yourstudentnumber where Yourstudentnumber = your student number.

Note: Marks will be deducted for submission in a non-PDF format.

Question 1. [5 marks] This question relates to the following table extracted from an article by Lima et al. entitled “Harmful drinking is associated with mental health conditions and other risk behaviours in Australian young people” published in the Australian and New Zealand Journal of Public Health (2020). Show your working and provide your answers to one decimal place.

- [1 mark] What percentage of young people who had a primary or secondary carer with a Bachelor degree or higher, had drunk alcohol in the last 30 days?
- [1 mark] Of all young people in the sample who were over the age of 14, what percentage had ever drunk alcohol?
- [1 mark] Calculate the mean age of young people who had drunk 4 or more drinks in the last 30 days. Comment on whether this is consistent with the overall average age of the sample.
- [2 marks] Calculate the median and inter-quartile range of age for all young people in this sample.

| | All young people | | Never drunk alcohol | | Ever drunk alcohol | | Drunk in last 30 days | | Drunk 4 or more drinks in the last 30 days | |
|------------------------------------------------------------------------|------------------|-----------------------------|---------------------|-------------------------|--------------------|-------------------------|-----------------------|-------------------------|--------------------------------------------|-------------------------|
| | Sample (N) | Population (%) ^a | Sample | Population ^a | Sample | Population ^a | Sample | Population ^a | Sample | Population ^a |
| Overall | 2,314 | 100.0% | 1,257 | 62.1% | 1,057 | 37.9% | 518 | 18.1% | 369 | 12.5% |
| Gender | | | | | | | | | | |
| Male | 1,192 | 51.3% | 658 | 52.0% | 534 | 50.0% | 266 | 51.0% | 198 | 51.0% |
| Female | 1,122 | 48.7% | 599 | 48.0% | 523 | 50.0% | 252 | 49.0% | 171 | 49.0% |
| Age (years) at interview^a | | | | | | | | | | |
| 13 | 310 | 19.7% | 284 | 29.0% | 26 | 4.8% | 5 | 2.3% | 2 | 1.6% |
| 14 | 343 | 20.0% | 268 | 26.0% | 75 | 11.0% | 25 | 8.2% | 14 | 6.8% |
| 15 | 309 | 19.7% | 207 | 22.0% | 102 | 16.0% | 48 | 15.0% | 30 | 13.0% |
| 16 | 718 | 20.2% | 305 | 14.0% | 413 | 31.0% | 193 | 30.0% | 148 | 34.0% |
| 17 | 634 | 20.4% | 193 | 9.9% | 441 | 38.0% | 247 | 44.0% | 175 | 45.0% |
| Socioeconomic status | | | | | | | | | | |
| 1 (High disadvantage) | 348 | 15.7% | 180 | 15.0% | 168 | 17.0% | 77 | 17.0% | 54 | 17.0% |
| 2 | 406 | 17.8% | 219 | 18.0% | 187 | 18.0% | 89 | 18.0% | 64 | 17.0% |
| 3 | 490 | 21.2% | 258 | 20.0% | 232 | 22.0% | 101 | 19.0% | 78 | 20.0% |
| 4 | 512 | 22.2% | 281 | 23.0% | 231 | 22.0% | 131 | 25.0% | 96 | 26.0% |
| 5 (Low disadvantage) | 558 | 23.1% | 319 | 24.0% | 239 | 21.0% | 120 | 21.0% | 77 | 20.0% |
| Family type | | | | | | | | | | |
| Original family | 1,468 | 63.7% | 868 | 69.0% | 600 | 56.0% | 288 | 53.0% | 202 | 52.0% |
| Step family | 140 | 5.7% | 67 | 5.2% | 73 | 6.6% | 35 | 6.4% | 28 | 7.5% |
| Blended family | 149 | 6.7% | 75 | 6.2% | 74 | 7.5% | 38 | 8.1% | 26 | 7.2% |
| Lone parent family | 529 | 22.8% | 236 | 19.0% | 293 | 29.0% | 151 | 32.0% | 109 | 33.0% |
| Other family | 28 | 1.1% | 11 | 0.8% | 17 | 1.5% | 6 | 1.0% | 4 | 0.9% |
| Highest level of education of either primary or secondary carer | | | | | | | | | | |
| Bachelor degree or higher | 887 | 38.0% | 524 | 41.0% | 363 | 33.0% | 187 | 36.0% | 142 | 38.0% |
| Diploma or certificate III/IV | 934 | 41.0% | 492 | 40.0% | 442 | 42.0% | 202 | 40.0% | 138 | 38.0% |
| Year 11 or 12 | 284 | 13.0% | 142 | 12.0% | 142 | 14.0% | 72 | 13.0% | 48 | 12.0% |
| Year 10 or below | 209 | 8.5% | 99 | 7.2% | 110 | 11.0% | 57 | 11.0% | 41 | 12.0% |

Notes:

^a: Percentage calculated from weighted sample

Percentage is calculated against the total number of children in each group.

Question 2. [2 marks] The table below provides baseline summaries from an article published in Australian and New Zealand Journal of Public Health by Grzeskowiak et al. in 2020, which looked at the effects of cannabis usage during pregnancy on neonatal outcomes.

- [1 mark] Write a single sentence describing the relationship between age and usage of cannabis before and during pregnancy.
- [1 mark] Which cannabis usage group had the highest coefficient of variation for Age? (Show your working and state the CV's in each group to three decimal places).

1 Characteristics of 5610 SCOPE study participants, 2004–2011, by cannabis use of mothers before and during pregnancy

| | Cannabis use | | | |
|-------------------------------------------------|--------------|-----------------------|-------------------------|---------------------------|
| | Never used | Quit before pregnancy | Quit early in pregnancy | Continued use at 15 weeks |
| Number of participants | 5296 | 97 | 157 | 60 |
| Age (years), mean (SD) | 28.9 (5.4) | 26.8 (5.9) | 24.6 (5.8) | 21.7 (4.9) |
| Body mass index (kg/m ²), mean (SD) | 25.3 (4.9) | 24.9 (4.5) | 25.5 (5.0) | 24.0 (5.5) |
| Socio-economic index, mean score (SD) | 42.3 (16.5) | 38.1 (15.9) | 32.4 (13.6) | 26.6 (9.0) |

Question 3. [2 marks] Consider the following Stem and leaf plot for SVAR for a sample of 231 patients admitted to the Emergency department.

SVAR Stem-and-Leaf Plot

```

Frequency      Stem & Leaf
1.00 Extremes      (= < 76)
1.00          7 . 9
4.00          8 . 0259
4.00          9 . 2257
16.00         10 . 0334444457888999
25.00         11 . 012222224455666777799999
46.00         12 . 0000001112222233444444455666666666666778888899
46.00         13 . 00000000111222233344444555566666666777778888999
28.00         14 . 0001111222233344455577778999
29.00         15 . 0000011222224444555677778888
15.00         16 . 011122233446678
6.00          17 . 566689
4.00          18 . 0046
6.00 Extremes      (>= 194)

```

Stem width: 10.00
Each leaf: 1 case(s)

- [1 mark] What is the 25th percentile for SVAR for this sample of patients? Show your working/reasoning.
- [1 mark] Provide a comment on what you believe the mean value will be relative to the median in this instance. Explain your reasoning.

Question 4. [11 marks] This question relates to the file **bsn81.sav**. See the doc on “Description of datasets” for information on this dataset. Use SPSS and the file **bsn81.sav** to produce the requested output and answer the following questions.

- a) [1 mark] Produce a histogram for the variable FVC and comment on the shape of the histogram.
- b) [2 marks] Group BMI to create a new variable called BMIGROUP with category labels, where the BMI groups are defined as:
Underweight: $\text{BMI} \leq 20.00$
Normal: $20.00 < \text{BMI} \leq 25.00$
Overweight: $25.00 < \text{BMI} \leq 30.00$
Obese: >30.00
Produce a single table that shows the count and percentage of people with asthma (ASTHMA=YES) by BMI group. Comment on the differences in these percentages across BMI groups.
- c) [2 marks] Produce side-by-side boxplots of FVC by ASTHMA and comment on the similarities and differences in the distribution of FVC for the two ASTHMA groupings.
- d) [2 marks] Produce a bar chart that shows the percent with asthma (ASTHMA=YES) by SEX and BMIGROUP (as in b). Compare the prevalence of asthma between males and females across BMI groups.
- e) [2 mark] Produce an error bar chart that shows the mean \pm 1 SD for FVC for each ASTHMA and BMIGROUP combination. Comment on the relationship between mean FVC and BMI separately for *those with asthma and those without asthma*.
- f) [2 marks] Obtain separately for males and females percentile estimates for FVC and highlight in your output the lower quartile, median and upper quartile values for FVC. What is the interval that contain the middle 80% of FVCs for males and females?