

## **Biostatistics DENT70001 2021/22: Resit Assignment (100%)**

**Submission deadline: 5th September, 2022; 1pm**

**Question 1 (40%) is for those who are required to resit the mid-assignment only.**

**Question 2 (60%) is for those who are required to resit the final-assignment only.**

### **[Question 1]**

**[40%]**

This question is based on a study regarding fluoride intake and dental fluorosis. The reference to the original paper can be found as below and can be downloaded from Blackboard.

Martins CC, Paiva SM, Lima-Arsati YB, Ramos-Jorge ML, Cury JA. (2008 ). "Prospective Study of the Association between Fluoride Intake and Dental Fluorosis in Permanent Teeth". Caries Res. 2008;42(2):125-33

### **1.1) Study Design**

**[24]**

1.1.1) Provide a brief description of the scientific question the authors aimed to answer. [2]

1.1.2) Apply the PECO-T (population (P), exposures (E) , comparison (C), outcome(s) (O), and time (T)) principle, to identify each of the P-E-C-O-T elements of the study. [10]

1.1.3) From a research terminology viewpoint, is it an experimental/observational, prospective/retrospective, cross-sectional/longitudinal study? Please select as appropriate and justify your statements. [6]

1.1.4) Suggest and describe in detail a different study design to that of the conducted study, to answer the same scientific question- which you have identified in (1.1.1)- that the authors aimed to answer. [6]

**[Please turn over]**

**1.2) INTERPRET RESULTS****[16]**

Use the main results in Table 2 from the paper (screen shot below) to answer the following questions:

**Table 2.** Median fluoride intake from diet, dentifrices and combined (mg F/kg body weight) by the children from the case (TFI  $\geq 1$ ) and noncase (TFI = 0) groups

| Source of fluoride | TFI = 0 (n = 20) |             | TFI $\geq 1$ (n = 29) |             | p <sup>1</sup> |
|--------------------|------------------|-------------|-----------------------|-------------|----------------|
|                    | median           | 95% CI      | median                | 95% CI      |                |
| Diet               | 0.029            | 0.026–0.038 | 0.031                 | 0.025–0.036 | 0.863          |
| Dentifrices        | 0.049            | 0.040–0.076 | 0.050                 | 0.040–0.068 | 0.943          |
| Combined           | 0.084            | 0.072–0.108 | 0.083                 | 0.068–0.101 | 0.669          |

<sup>1</sup> Mann-Whitney test.

1.2.1) Identify the type of variable (e.g. continuous/categorical/ordinal/binary) that “fluoride intake” is. What is the likely reason that the authors reported “Median” rather than “Mean” fluoride intake? [4]

1.2.2) Identify the statistical method the authors used to calculate the P-value (last column in the Table 2). What could be the reasons to support authors’ choice of using this method? [4]

1.2.3) Using the results from the third row (labelled as “Combined”), state clearly the Null and Alternative hypotheses the authors were testing. Using the values provided to write a short report and conclusions based on the results of this third row. [8]

**[THE END OF QUESTION 1]**

**[Please turn over]**

## [Question 2] [60%]

Question 2 is a computer-based exercise using the data we have collected from the **second** face-to-face (F2F) session. The original data have been modified to suit the purpose of the current assignment. As stated at the data collection session, this survey is anonymised. The data are only for teaching and assessment purposes within this course and will not be distributed anywhere else.

Students who came to the de-briefing sessions had been asked the following 5 questions on two occasions. In this assignment we will analyse the data from the second F2F session. These questions are intended to determine Manchester Dentistry-PG students' perception of statistics. All PG students registered for this year's course should have an entry in the dataset; however, not everybody filled in the questionnaire. The questionnaires have been coded into a dataset by the following rules ("**SecondSurvey2022.Resit.BB.csv**"). For those who did not provide answers, a missing entry coded as "NA" was created.

| Variable Name          | Questions  | Coding/unit  |
|------------------------|--|--|
| <b>ID</b>              | ID   | Randomly assigned  |
| <b>Gender3</b>         | <b>2<sup>nd</sup> F2F:</b> Gender recorded at 2 <sup>nd</sup> session  | Female=1; Male=0; Missing=NA   |
| <b>LikeStat3</b>       | <b>2<sup>nd</sup> F2F:</b> I like statistics   | 1=Strongly agree; 2= agree;<br>3=neutral; 4=disagree;<br>5=Strongly disagree |
| <b>ConfidenceStat3</b> | <b>2<sup>nd</sup> F2F:</b> I am confident in using statistics<br>(provide a score between 1 and 100; with 1=no confidence and 100=very confident. Can provide a figure with decimal place) | 1 to 100   |
| <b>Exercise3</b>       | <b>2<sup>nd</sup> F2F:</b> The time I spend on physical exercise per week (Minutes)  | (Minutes)<br>Missing=NA  |
| <b>UseStatJob3</b>     | <b>2<sup>nd</sup> F2F:</b><br>I am likely to use statistics in my career   | Yes=1; No=0; Missing=NA  |

You are allowed to use any statistical software (or calculator) familiar to you to complete the exercise. While requiring you to demonstrate evidence of computer competency, you should focus more on correctly presenting and interpreting the data/analysis/results. The dataset "**SecondSurvey2022.Resit.BB.csv**" can be downloaded from Blackboard.

2.1) How many students have been enrolled into the course? How many students had responded to the questionnaire? How many non-responders are there in the dataset? [3]

[Please turn over]

- 2.2) It is important to examine each of the variables before conducting any analysis. [16]
- 2.2.1) Identify any potential errors in the dataset. [8]
- 2.2.2) Propose solutions and apply them to your dataset to fix the issues. [4]
- 2.2.3) Once any issue has been fixed in (2.2.2), you will use this updated dataset to answer the rest of the questions in this section. State clearly how many students' records are there in your updated dataset? How many variables are there in this dataset? [4]
- 2.3) We are interested in studying if there are differences in the exercise duration between the two genders by using the variables **"Gender3"** and **"Exercise3"**. [41]
- 2.3.1) Identify the data type of the variable **"Exercise3"** (e.g. Qualitative/Quantitative; Discrete (categorical)/Continuous/Binary/ordinal/etc.). [4]
- 2.3.2) Identify the data type of the variable **"Gender3"** (e.g. Qualitative/Quantitative; Discrete (categorical)/Continuous/Binary/ordinal/etc.). [4]
- 2.3.3) Calculate summary/descriptive statistics for exercise duration for each gender (variable **"Exercise3"** and **"Gender3"**). Your answer should include the following information: n (number of students), minimum, maximum, median, IQR, mean, and standard deviation. [7]
- 2.3.4) Generate two box and whisker plots for **"Exercise3"**, one for female and one for male. Include the boxplots in your answer. Your figures should have a meaningful title and be clearly labelled. Briefly describe different aspects of the box and whisker plot (e.g. how the size of the box and the length of the whisker are decided, and explain the key components that appear on your figure). [10]
- 2.3.5) Using the numbers and plots from questions (2.3.3) and (2.3.4), describe the distributions of the variable **"Exercise3"** between the two different genders (**"Gender3"**). Write a paragraph (no more than 200 words) to summarize your findings related to the students' weekly exercise duration and genders. [6]
- 2.3.6) What is the average exercise duration in a week for female students? What is the average exercise duration in a week for male students? What is the difference in the period of exercise on average between the two genders? [3]
- 2.3.7) Conduct an analysis to study the association between weekly exercise duration and gender using the variables **"Exercise3"** and **"Gender3"**. State clearly the hypothesis, statistical test and results (your results should consist of the point estimates, test statistics, P-value, and confidence interval) of the test, and draw your own conclusions. [7]

[THE END OF QUESTION 2]