



THE UNIVERSITY OF
SYDNEY

Assignment

PUBH5018 Introductory Biostatistics

Due Date and Time: Monday, March 30 2020 11:59 PM Sydney local time

Assignment Category: Submitted Work

Assignment Sub-category: Assignment

Plagiarism Policy

You must complete your assignment alone. Submitting assignments which have been jointly done is not acceptable. Copying someone else's work or quoting from text without adequate attribution of the source is plagiarism and is not acceptable. All assignments will be verified by plagiarism detection software. Penalties will be applied for plagiarism. The University's policy on academic honesty can be found at the following site:

<http://sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2012/254&RendNum=0>

Late Penalties & Special Consideration

Unless you have made a prior arrangement with the Unit Coordinator, have special consideration or an academic plan, 1 mark (5% of 20) will be deducted from your assignment mark per day (or part thereof) until Thursday, April 9, 11:59 PM Sydney local time. Assignments submitted past this date without approved special consideration or an academic plan will not be accepted and will be given a zero (0) mark. For students seeking 'Special consideration' please use the following site:

http://sydney.edu.au/current_students/special_consideration/apply.shtml

Instructions to Students

The data for the assignment has been derived from the online Introductory Biostatistics Questionnaire that students completed in the first week of semester. We had 312 responses – thank you! For this assignment each student has been allocated their own data set, comprising of a random sample of respondents. A few modifications to the data have been made to make this data suitable for the Assignment.

Datasets are named BIOQAssig###.omv, where ### is a 3-digit number which is listed against your name in the **Assignment Dataset Allocation** file. For example, if your allocated number is 012, then your data set is BIOQAssig012.omv. Please ensure you use your allocated dataset. All datasets are available within the **PUBH5018 Assignment Datasets** folder, under Assessment Resources.

Important Notes

1. This assignment paper (including these instructions) is **five (5) pages** in length. Please ensure you have all pages.
2. Please ensure you **use your allocated data set**. Because the datasets differ between students, the results will differ, and the conclusions may also vary.
3. The **variable names and coding** of the variables in your dataset are attached at the end of this assignment on page 5.
4. There is **not always just one correct way of handling data**: you are sometimes required to use your judgment. When this occurs, you should **justify the decision** you have made.
5. Submit your assignment in a **SINGLE** Word or PDF document.
6. **jamovi output presented must be edited** when needed and possible to comply with the guidelines given in this course.
7. **Name your file submission** with your student number (SID), unit code and BIOQAssig dataset number e.g. 311275249_PUBH5018_BIOQAssig026.docx. Please number all pages. Also show page numbers, BIOQAssig dataset number and your student number in the header or footer of your assignment.
8. Do **NOT** put your name ANYWHERE on your assignment or the submission title.

*If you have any administrative questions, please post them on the **CANVAS discussion board**.*

*Go to **Discussions > Assignment Discussion – General Queries**.*

Alternatively contact Thomas Demmery sph.epibio@sydney.edu.au

*If you **have difficulties submitting the assignment** around the due time, please email*

*sph.epibio@sydney.edu.au directly **with your assignment attached** to avoid late penalties. The time stamp of your email will be used as evidence of the date and time of your assignment submission. Please note responses to emails will only occur on standard working days.*

Assignment Questions

In this assignment you will be analysing the data collected in the class survey. A copy of this survey is found on the last page of this assignment.

Note: All tables and graphs included in your answers for questions 1 through 7 must conform to the PUBH5018 presentation guidelines. Penalties will be applied where they do not conform.

Questions

1. (2.5 marks)

Using appropriate methods, individually examine the following variables: *lang*, *eyewear*, *sex*, *age*, *employ*, *breath* and *sitting* to determine if there are any missing or outlier values. If necessary, delete any values as appropriate. For each variable, comment very briefly on what analysis you used, report what you found, any data edits you performed (if any) and why. (Preferably present it as a table; for example, you could use the template below); otherwise it must be a **maximum of one sentence** per variable).

NB: You should use your cleaned data, i.e. the data with any edits performed, for the remainder of the assignment.

Table 1: Analysis of variables

Variable	Analysis used	Number missing	Outlier/Implausible values description	Action taken description
lang				
eyewear				
sex				
age				
employ				
breath				
sitting				

2. (3 marks)

Calculate a statistical anxiety score (SAS) for each participant using the formula:

$$SAS = 10 + anx1 + anx2 - \frac{anxi3^2}{10}$$

Display the distribution of SAS using a boxplot and describe the important features of this distribution with relevant summary statistics.

3. (2 marks)

Long sitting times every day are associated with poorer health outcomes. Group the variable *sitting* into three categories as follows: Low risk (less than 4 hours sitting); medium risk (at least 4 hours sitting, but less than 8 hours); high risk (at least 8 hours sitting). Produce a frequency table that includes the number and percentage of participants in each category.

4. (2 marks)

Produce the appropriate plot that shows a breakdown of eye wear (*eyewear*) used by participants.

5. (2 marks)

Produce a histogram of age (*age*). Write one to three sentences that describe the distribution and provide appropriate summary statistics.

6. (3 marks)

Create a two-way table with *lang* as the rows and *employ* as the columns. Within each employment category, include the percentage of participants that do and do not speak a language other than English at home. Summarise the table in 1 to 2 sentences.

7. (5 marks)

Perform a hypothesis test that investigates whether Introductory Biostatistics students have a mean breath holding time (*breath*) of 36 seconds. To this end, carry out the following steps:

- State the null hypothesis for this hypothesis test
- Carry out a one sample t-test
- Calculate a 95% confidence interval for the mean breath holding time (**Note: this is not the same as the confidence interval for the mean difference**)
- Write a conclusion (2 – 4 sentences long) that summarises the important results

8. (0.5 marks)

To receive the marks for this question, you need to have created your profile with Examity through Canvas by the assignment due date. To do this, follow the instructions below:

- Log onto the PUBH5018 Canvas site
- Navigate to **Modules > Examity Mid-Semester Online Assessment 2020**
- Read through the “Examity Registration Instructions for Students - 2020” document
- Create your profile with Examity following the instructions in the above document

Total = 20 marks

This is the end of the assignment questions

The **variable names and survey questions** used for the data are on the following page

Variable names and coding for the Introductory Biostatistics Questionnaire data 2020

Survey question	Variable name
1. What degree program are you enrolled in?	degree
2. Do you speak a language other than English at home (yes/no)	lang
3. What corrective vision eyewear do you use (glasses, contacts, both, monocle, none)?	eyewear
4. What is your Sex (Male, Female, Intersex)	sex
5. What is your age in years?	age
6. Approximately how much money in dollars did you spent in the previous week on groceries?	groc
7. How long (in seconds) can you hold your breath?	breath
8. What is your employment status (Full-time, Part-time, Not employed)	employ
9. When did you last visit a general practitioner? (Less than 7 days ago, Greater than 7 days, but less than 1 month ago, Greater than 1 month, but less than 6 months ago, Greater than 6 months ago, but less than 1 year ago, Greater than 1 year ago, but less than 2 years ago, Greater than 2 years ago, Never)	doctor
10. About how many hours (to the nearest 0.25 hours (15 minutes)) did you spend sitting down yesterday?	sitting
11. On a scale from 0 (not at all anxious) to 10 (extremely anxious), how anxious are you about learning biostatistics?	anxi1
12. On a scale from 0 (not at all anxious) to 10 (extremely anxious), how anxious are you about using mathematical equations?	anxi2
13. On a scale from 0 (not at all confident) to 10 (extremely confident), how confident are you to learn new computer software?	anxi3