Northumbria University NEWCASTLE

ASSESSMENT BRIEF

Module Title:	Statistics and Business Intelligence
Module Code:	PE7050
Academic Year / Semester:	2022-23 / Semester 2
Module Tutors / Email (all queries):	Dr Naveed Anwar Naveed.Anwar@northumbria.ac.uk
% Weighting (to overall module):	100%
Assessment Title:	Statistics and Business Intelligence Questions
Date of Handout to Students:	9th January 2023
Mechanism for Handout:	Module Blackboard Site
Deadline for Attempt Submission by Students:	26th March 2023 11.59pm GMT
Mechanism for Submission:	Document to upload to Module Blackboard Site
Submission Format / Word Count	Please upload your solution to questions (problems) as a single PDF/Word document.
Date by which Work, Feedback and Marks will be returned:	21st April 2023
Mechanism for return of Feedback and Marks:	Marks with feedback will be uploaded to the Module Site on Blackboard. For further queries please email module tutor.
Further instructions:	• This is an individual assessment. • You are advised to be familiar with the University's definitions of academic misconduct including plagiarism and collusion (Northumbria, 2018a). • You are required to use the Harvard Style of referencing and citation. The "Citethem right" guide is recommended for referencing and citation (Pears and Shields, 2008) which should be followed throughout your assignment. A good alternative is Northumbria (2018b). • ***Students are required to use RStudio software (or any other software allowed by the tutor) and attach its code (instructions) and screenshots (showing the details of the work done) to their solution.*** • ***Students are required to provide the details how an answer is achieved - all the calculations / logics / instructions / code are to be provided/discussed.*** • The number of marks allocated to each of the questions are denoted in parentheses "( )".

References

- Northumbria (2018a) Academic Regulations for Taught Awards. Available at: https://northumbria-cdn.azureedge.net/-/media/corporate-website/new-sitecore-gallery/services/academic-registry/documents/qte/assessment/guidance-for-students/pl049-v002-academic-regulations-for-taught-awards-2021-22-for-levels-3-4-5_new-l7.pdf?modified=20210910145131 [Last accessed: 11th January 2022]
- Northumbria (2018b) Quick guide to Referencing and Plagiarism. Available at: https://cragside.northumbria.ac.uk/Everyone/skillsplus/database_uploads/87.pdf [Last accessed: 11th January 2022]
- Pears, R. and Richard, S, and Shields, G. J. (2019) Cite them right: the essential referencing guide. Eleventh Revised and Expanded ed. 2019. Print. Macmillan Study Skills. Available at: https://silo.tips/download/cite-them-right-the-essential-referencing-guide-richard-pears-and-graham-shields [Last accessed: 11th January 2022]

Learning outcomes:

Knowledge & Understanding:

1. Key aspects of statistics and business intelligence techniques for forecasting future trends and business growth
2. Application of appropriate techniques for understanding and exploring business data and how it can be exploited for decision making

Intellectual / Professional skills & abilities:

3. Apply different statistics and business intelligence techniques to complex real-world business environment having to deal with and exploit huge volumes of data
4. Critically evaluate and reflect on solving real life business problems using suitable exploratory data analysis techniques and algorithm

Personal Values Attributes (Global / Cultural awareness, Ethics, Curiosity) (PVA):

5. Build an awareness of legal, cultural and ethical issues surrounding analysis, exploration and dissemination of data

This assessment addresses Learning Outcomes 1, 2, 3, 4 and 5.

Assessment Questions

1. Six months ago, a local gym set up a research programme to find out if gym members who attended exercise classes were more likely to lose weight than those who exercised alone. A census of all participants was conducted. These were the results they recorded:

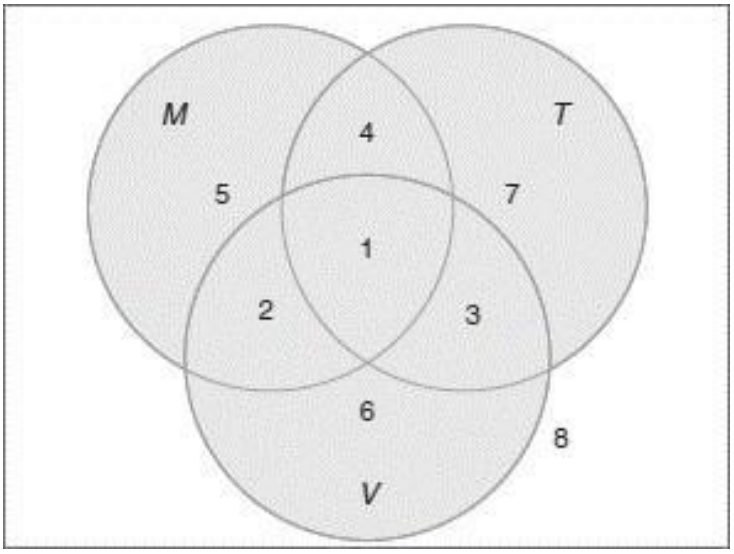
	Exercise class	Gym-only workouts
Participants	45	62
Mean weight loss over 6 month	1.6 kgs	2.3 kgs
Mode weight loss over 6 months	1.3 kgs	1.5 kgs
Standard deviation	1.03	1.33

The staff at the gym want to know which type of exercise – gym only workouts or attending exercise classes – is most effective in helping individuals lose weight. Prepare a short report (not more than a page or two) which summarises and interprets the findings, using all of the statistics given in the table above. **(11)**

2. Describe a way to deal with missing data values in data for processing it. [No more than 500 words] **(5)**

3. Suppose that a family is leaving on a summer vacation in their camper and that M is the event that they will experience mechanical problems, T is the event that they will receive a ticket for committing a traffic violation, and V is the event that they will arrive at a campsite with no vacancies. Referring to the Venn diagram of this situation in the Figure below, state in words the events represented by the following regions: **(10)**

(a) region 5
(b) region 3
(c) regions 1 and 2 together
(d) regions 4 and 7 together
(e) regions 3, 6, 7, and 8 together



4. According to Consumer Digest Magazine the probable location of an LCD TV in the home is as follows: **(6)**

Adult bedroom: 0.15
Child bedroom: 0.07
Other bedroom: 0.12
Office or den: 0.38
Other rooms: 0.28

(a) What is the probability that an LCD TV is in a bedroom? (2)
(b) What is the probability that it is not in a bedroom? (2)
(c) Suppose a household is selected at random from households with an LCD TV; in what room would you expect to find an LCD TV? (2)

5. The probability that an iPhone will survive a shock test is 0.70. Find the probability that exactly 3 of the next 5 iPhones tested survive. These tests are independent. **(6)**

6. A real estate agent claims that 63% of all private residences being built today are 3-bedroom homes. To test this claim, a large sample of new residences is inspected; the proportion of these homes with 3 bedrooms is recorded and used as the test statistic. State the null and alternative hypotheses to be 3 in this test and determine the location of the critical region. **(6)**

7. The following data is taken from a company about its advertisements and purchases of the product. Calculate coefficient of correlation to measure the strength and direction of relationship between the number of advertisements and purchases made, and comment on it. **(6)**

Number of advertisements	10	7	6	5	4	3	2	0
Purchases	12	3	8	10	5	4	1	4

8. A famous company selling household appliances wants to determine the relationship between advertising expenditures and sales. The following data was taken from 6 major sales regions. The expenditure is in thousands of pounds and sales are in millions of pounds. **(8)**

Region	Expenditure, x	Sales, y
1	1.5	2.0
2	4.5	3.0
3	8.0	4.5
4	4.0	2.5
5	2.0	2.0
6	4.0	5.0

(a) Estimate the linear regression line to provide a chart and summary statistics together with the coefficients. (6)
(b) Estimate the expected sales for a region where 6.3 thousand pounds are being spent on advertising. (2)

9. Below are given some hearing frequencies (audiograms), you are required to **(12)**
(a) Find the number of clusters for the given data (4)
(b) Cluster the given data and comment on each cluster of the data (8)

[Hint: Read this paper – Anwar, Naveed, Oakes, Michael, Wermter, Stefan and Heinrich, Stefan (2010) Clustering audiology data. In: 19th Annual Machine Learning Conference of Belgium and TheNetherlands, 27-28 May 2010, Leuven. Web link: https://dtai.cs.kuleuven.be/events/Benelearn2010/submissions/benelearn2010_submission_7.pdf]

Freq250	Freq500	Freq1K	Freq2K	Freq4K	Freq8K
50	60	60	60	70	70
50	55	65	70	75	130
15	15	25	60	70	55
75	75	75	70	75	70
20	15	15	55	65	75
70	65	75	70	65	80
20	20	40	45	50	60
30	25	20	30	45	55
30	30	40	55	75	95
45	45	45	65	80	70
30	20	25	35	70	80
15	15	25	55	90	95
50	50	55	50	50	75
30	30	25	55	70	130
35	50	45	60	80	130
55	60	70	60	95	90
40	30	30	30	70	85
30	30	40	35	55	65
35	25	35	35	55	75
35	35	50	50	80	100
65	65	65	65	70	70
40	40	35	50	60	65
20	20	40	50	85	130
35	35	35	55	60	75
25	20	20	65	70	55
50	45	40	25	30	70
55	45	50	40	35	35
45	35	40	55	80	130
15	15	10	55	65	70
50	55	60	60	70	85
35	35	50	55	70	85
65	60	55	55	60	65
60	50	35	50	60	95
20	20	10	30	75	75
20	30	30	45	65	80
35	30	30	60	60	65
30	45	50	50	50	80
50	65	55	50	50	80

80	60	45	45	65	130
30	25	30	35	50	75
70	70	70	60	85	85
55	55	55	55	65	130
40	35	45	45	55	65
45	40	40	50	65	85
60	75	65	65	75	130
20	10	15	55	60	70
35	50	65	65	70	95
65	65	65	65	70	75
60	55	30	40	65	70
30	30	40	30	70	55

10. A scenario **(30)**

You are a director of a major manufacturing organisation and collecting various pieces of information for your potential clients, such as on one of your major clients who is based in London, will require delivery lorries to travel the length of the M1. You will investigate the speed on this road using the data available at https://www.trafficengland.com/traffic-report .

You should only use the source specified. You will need to adopt a sampling approach and credit will be given for schemes which show you have considered how to apply the principles of sampling to obtain the best results with the smallest possible dataset.

Report Requirements
Your report should consist of eight or fewer pages (applies only to question 10) and should be word-processed. Credit will be given for the use of an appropriate technical style of presentation.

Your report should address the following topics:
(a) Your sampling strategy and how it was devised
(b) Details of the data collected
(c) Details of your statistical analysis and its results
(d) Conclusions drawn
(e) Any relevant background research

Credit will be given for an appropriate use of graphs, tables and charts. All external sources of information must be correctly cited and referenced.

You should include a table of all the data you have collected, and any calculations performed in RStudio as an appendix to your report. This is not counted in the page limit. Failure to include this will result in the deduction of marks. Also note that the Traffic England website (link given above) allows you to sign in and save data shots, but any updates on their website may lead to losing of data stored in your account. Thus, it is your responsibility to store collected data into your computer(s) and keep it safe for your above task.

Mark Scheme for this question is given below in the table:

Marks range in Percentage	***First (70 - 100)***	***Upper Second (60 - 69)***	***Lower Second (50 - 59)***	***Third (40 - 49)***	***Fail (30 - 39)***	***Fail (1-29)***	**Tutor Comments & Feedback**
General Criteria → ***Assessment criteria*** ↓	*Knowledge and understanding is comprehensive & exceptional both as to breadth and depth.*	*Knowledge base is up-to-date and relevant, but also may be broad or deep.*	*Sound understanding of topic. Reasoning and argument are generally relevant but not necessarily extensive.*	*Knowledge is adequate but limited and/or superficial. Description/assertion rather than argument. Insufficient focus.*	*Knowledge is limited and/or superficial; frequently inarticulate and/or irrelevant.*	*Minimal awareness of subject area. Work is inarticulate and/or incomprehensible.*	
Data Collection **20%**	Excellent - description of sample strategy, consideration of bias and presentation of raw data in appendix	Very Good - description of sample strategy, consideration of bias and presentation of raw data in appendix	Good - description of sample strategy, consideration of bias and presentation of raw data in appendix	An adequate - description of sample strategy, consideration of bias and presentation of raw data in appendix	A brief - description of sample strategy, consideration of bias, and presentation of raw data in appendix	Inadequate - description of sample strategy, consideration of bias and presentation of raw data in appendix	
Data analysis 34%	Excellent - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	Very Good - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	Good - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	An adequate - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	A brief - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	Inadequate - selection of appropriate descriptive statistics, description of the techniques used, application of statistical analysis, presentation of results	

Conclusion(s) 23%	Excellent conclusions - reflect the results of the analysis, conclusions are relevant to the original question	Very Good conclusions - reflect the results of the analysis, conclusions are relevant to the original question	Good conclusions - reflect the results of the analysis, conclusions are relevant to the original question	Adequate conclusions - reflect the results of the analysis, conclusions are relevant to the original question	Brief conclusions - reflect the results of the analysis, conclusions are relevant to the original question	Inadequate conclusions - reflect the results of the analysis, conclusions are relevant to the original question	
Report 23%	Excellent report - containing references, figures, style, presentation	V good report - containing references, figures, style, presentation	Good report - containing references, figures, style, presentation	Adequate report - containing references, figures, style, presentation	Brief report - containing references, figures, style, presentation	Inadequate report - containing references, figures, style, presentation	