Report

Due Date: 20 October 2022, 11.59 pm AEST
Weighting: 40%
Full Marks: 100

- Please read this document fully and carefully before starting your analyses.
- There are two parts to this assessment.
- Part A is a report of project analysis of data as described in detail on pages 3-5 of this document. Your submission will be no more than 5 pages for Part A.
- Part B is the completion of a set of 10 Multiple Choice questions and 5 Short Answer questions on page 6 - 9 of this document.
- Part A and B should be submitted together in one pdf file to the link provided on the StudyDesk.
- Make sure all work and wording are your own. Submissions will be checked for potential plagiarism.
- Remember that your job in this assessment is to demonstrate your understanding of the dataset, the analysis method, and the interpretation of results. This does not mean you will need to do any additional literature search on the NBA or basketball to describe the data or interpret your results.
- Plots and tables from SPSS can be copied directly from SPSS into your report, but you should consider whether all output is needed or relevant.
- Remember to describe/interpret all statistics, plots, and results of analysis that you include in your report.
- This assessment assesses the following course objectives:
  - ➢ Distinguish between different methods of data generation, collection and analysis.
  - ➢ Select and implement appropriate statistical methods to analyse quantitative data to answer an underlying research question.
  - ➢ Apply professional skills to effectively present and communicate results of statistical analyses including interpretation and justification of statistical decisions to a diverse scholarly audience.

Marks distribution:

Part A: 60 marks total

Introduction: 6 marks

Methods: 20 marks

Results: 30 marks

Conclusions: 4 marks

Part B: 35 marks total

Question 1:

10 Multiple Choice questions. 2 marks each. *20 marks total*

Question 2:

5 Short Answer questions. 3, 3, 2, 3, 4 marks respectively. *15 marks total*

Presentation of Report (Part A) and Part B (including spelling and grammar)**: 5 marks**

Part A

1.0   Objective:

The purpose of this project is to provide you with an opportunity to demonstrate synthesis, understanding and communication of some of the concepts, statistical methods, and practical analyses that you have learnt in this course. **This project focuses on the multiple regression method including:**

1. Creating and defining an SPSS data set.
2. Data exploration, descriptive analysis and fitting an appropriate linear model.
3. Interpreting the results produced by SPSS.
4. Checking the assumptions of the fitted model.
5. Preparing/writing a report on the project as a research report within 5 pages (maximum).

2.0   The Data:

A consultancy company (the client) has asked you to explore some data about basketball players. The dataset, **Basketball.xlsx**, is a random sample of major league basketball players (N=54 players). The following variables were recorded for each player:

X1 = height in feet
X2 = weight in pounds
X3 = percent of successful field goals (out of 100 attempted)
X4 = percent of successful free throws (out of 100 attempted)
X5 = average points scored per game
Reference: The Official NBA Basketball Encyclopedia, Villard Books

The dataset is available on the assessment page of the STA8170 StudyDesk. Import this dataset into SPSS (no hand calculations are required). If you have any difficulty importing the dataset, please first check the relevant video under the SPSS Video Demonstrations link on the StudyDesk. Make sure all sections of the Variable View are completed correctly before you begin addressing the Tasks below.

First, perform Tasks 1 and 2 to address the Research Question defined below. Once you have completed all your analysis (Section 4.0) and are happy with your results then write a 5-page (maximum) report for the client (consultancy company). The structure of your report should follow the report template provided on in Section 5.0 (pages 4-5).

Note: *Even though you are to work through Sections 3.0 and 4.0 these do not constitute your Part A submission. You need to incorporate these responses within the 5-page report (as per Section 5.0, which will use some of your responses from Section 4.0).*

3.0   The Research Question:

The client would like to know: Can the average points scored per game be effectively modelled by the height and weight of players and their percent of successful field goals and free throws?

To address this question, you will need to consider the complete data analysis process as outlined in Section 4.0 below, and then communicate your process and findings as described in Section 5.0 below.

4.0   Data Analysis Process:

   a. Identify the appropriate response and explanatory variables based on the information provided in Sections 1.0, 2.0 and 3.0 above.
   b. Select appropriate summary statistics to describe: the dataset, each of the variables, and the relationships between pairs of variables. You may wish to include plot(s) to help you describe linear relationships between variables.
   c. State and define the appropriate multiple regression model and null hypothesis for the model. Perform the multiple regression analysis and interpret the overall significance of the model.
   d. Interpret the significance of all individual regression parameters. Find and interpret the 95% confidence interval (using SPSS, not by hand) for the most significant regression coefficient. State the fitted regression model.
   e. Discuss the assumptions of the analysis method in context.
   f. Discuss the overall suitability of the regression model, reporting appropriate statistics as part of your discussion.

5.0 Template for the Report (5 pages maximum)

Once you have competed the Data Analysis Process (Section 4.0), you will then need to communicate your work in the form of a Report (5 pages maximum). This report will form your submission for Part A. The report should be written using any technical terms needed to convey information correctly, but also needs to be understandable to the client (assume the client has a workable statistical knowledge).

General guidelines:
   • There are no marks given for referencing, however you are encouraged to attempt this as per your discipline if you have learnt how to do so (the reference list isn't included in your 5-page limit).
   • Note that all reports will be checked for plagiarism, so make sure it is your own work.
   • You should only include relevant SPSS output in your report.
   • Do not include an Appendix.
   • Submit only 1 pdf file which includes your Part A Report and Answers to Part B.

Introduction (7 marks)

- This should include a brief introduction into your report (suggest 0.5 to 1 page), specifically
  - The purpose of the overall dataset and a brief overview of the dataset
  - What data/variables you are analysing in the report
  - What questions you are going to address in the report, ensuring you cover **4a in the Data Analysis Process** items.
  - Try to keep this section succinct so that you have more room for the analysis and interpretation (see sections below).
  - You are not expected to do any additional research on the data context – all the information required for the Introduction is provided to you in this document.

Methods (20 marks)

- This section requires a clear and concise description of what and how the analysis was performed, and the rationale for why specific procedures were chosen. This includes methods relevant to **4b – 4f inclusive of the Data Analysis Process** items (suggest 1 - 1.5 pages).
- This section should not include any results, just describe what you did and why.

Results & Interpretation (30 marks)

- This will be the largest section and includes all your Tables, Figures, and results from your analyses results (suggest 1.5 – 2.5 pages).
- Clearly present and interpret the results from all methods described in the methods section (i.e. **4b – 4f inclusive of the Data Analysis Process** items).
- You can use tables and figures (plots) to help you describe the results. You can copy/paste figures and tables directly from SPSS, however only relevant output that you interpret in context should be included. (Remember: if you are using SPSS via Turbo you will need to take screen shots of output rather than copy/paste.)

Conclusions (13 marks)

- Provide a short, final conclusion answering the research question asked by the client (suggest <0.5 pages)

******************* End of Part A *******************

Part B

Question 1: 10 multiple choice questions. In your document for submission, list Q1.1 to Q1.10 and identify your chosen answer number, as per the table below (i.e., if you think the correct answer for Q1.1 is answer 2 then type only **Q1.1 (2)** in your submission document). Do not reproduce all the question text in your document. (*20 marks total*)

Question	Answer
Q1.1	
Q1.2	
Q1.3	
Q1.4	
Q1.5	
Q1.6	
Q1.7	
Q1.8	
Q1.9	
Q1.10	

Submit the above table, with your answers only for this section.

Question 1.1: (*2 marks*)

Which of the following does not affect the width of a confidence interval for a mean?

(1) The sample size.

(2) The level of confidence.

(3) The population size.

(4) The variability of the data.

(5) The standard deviation of the data.

Question 1.2: (*2 marks*)

Suppose in a simple regression, the slope b = -1. What can we conclude?

(1) r is negative.

(2) r^2 = 1.

(3) X and Y are significantly correlated.

(4) X and Y have equal means but of opposite signs.

(5) $r > 0$.

Question 1.3: (*2 marks*)

If you took a large random sample, you would expect the standard deviation of that sample to

be:

(1) Close to μ.

(2) Near in value to the population standard deviation.

(3) Smaller than that for a small random sample.

(4) More than the population standard deviation.

(5) Less than the population standard deviation.

Question 1.4: (*2 marks*)

A student survey was conducted with on-campus students to study the relationship between

place of residence (house, apartment, residential college, parent's home) and how the students

travel to campus (walk, bicycle, car, bus, train). What are the degrees of freedom for the Chi-

square statistic?

(1) 3

(2) 4

(3) 8

(4) 12

(5) 20

Question 1.5: (*2 marks*)

What type of graph best shows the relationship between a categorical variable and a

quantitative variable?

(1) Scatterplot

(2) Pie chart

(3) Side-by-side boxplot

(4) Histogram

(5) Bar chart

Question 1.6: (*2 marks*)

You obtain a sample chi-square value of $\chi^2 = -2.57$. Which of the following statements is true?

(1) There is a negative association between the two variables.

(2) There is calculation error, the χ^2 value cannot be negative.

(3) There is no association between the two variables.

(4) The observed frequencies are higher than the expected frequencies across at least half of the
cells.

(5) The observed frequencies are lower than the expected frequencies across all cells.

Question 1.7: (*2 marks*)

When calculating a confidence interval for proportion using $\hat{p} \pm z \times SE(\hat{p})$, which of the

following is not one of the assumptions made?

(1) $n\hat{p} > 10$

(2) $n(1 - \hat{p}) \geq 10$

(3) The sample size is smaller than 10% of the population.

(4) The sample is randomly chosen from the population.

(5) The sample size is greater than 30.

Question 1.8: (*2 marks*)

Given a frequency distribution is positively skewed (skewed to the right), which one of the

following statements is correct?

(1) The mean is less than the median.

(2) The mean and median are the same.

(3) The median is affected by skewness.

(4) The distribution has more large values than small values.

(5) The mean is greater than the median.

Question 1.9: (*2 marks*)

Which of the following is an example of a matched pairs design?

(1) A teacher compares the pre-test and post-test scores of students.

(2) A teacher compares the scores of students using a computer-based method of instruction
with the scores of another group of students using a traditional method of instruction.

(3) A teacher compares the scores of students in her class on a standardized test with the
national average.

(4) A teacher calculates the average score of students on a pair of tests and wishes to see if this
average is larger than 80%.

(5) None of the above.

Question 1.10: (*2 marks*)

The members of a health club were divided into three different age groups and a random sample taken from each of the groups. What type of sampling technique is this?

(1) Cluster sampling.

(2) SRS.

(3) Systematic sampling.

(4) Stratified sampling.

(5) Convenience.

Question 2: 5 short answer questions. In your submission document clearly label each of your answers for Question 2.1 to Question 2.5. Do not reproduce the question text in your document. Note that all responses will be checked for plagiarism so make sure it is your own work (*15 marks total*)

Question 2.1: (*3 marks*)

Clearly explain cluster sampling and give an example of when you might do this type of sampling (70 words max)

Question 2.2: (3 *marks*)

Clearly explain the Type 1 error in the context of a hypothesis test (100 words max)

Question 2.3: (*2 mark*)

What is the difference between a parameter and a statistic? (50 words max)

Question 2.4: (3 *mark*)

What is the purpose of standardising data? (100 words max)

Question 2.5: (*4 mark*)

What is the difference between an experimental and an observational study? Give an example of each to clearly define this difference (120 words max)

****************** **End of Part B** ********************

****************** **End of Assessment** ********************