

BCB744 (BioStatistics): Final Integrative Assessment

Smit, A. J.

2024-04-10

Honesty Pledge

This assignment requires that you work as an individual and not share your code, results, or discussion with your peers. Penalties and disciplinary action will apply if you are found cheating.

Acknowledgement of the Pledge

Copy the statement, below, into your document and replace the underscores with your name acknowledging adherence to the UWC's Honesty Pledge.

I, _____, hereby state that I have not communicated with or gained information in any way from my peers and that all work is my own.

Instructions

Please note the following instructions. Failing to comply with them in full will result in a loss of marks.

- **QUARTO → HTML** Submit your assessment answers as an .html file compiled from your Quarto document. Produce *fully annotated reports*, including the meta-information at the top (name, date, purpose, etc.). Provide ample commentary explaining the purpose of the various tests/sections as necessary.
- **TESTING OF ASSUMPTIONS** For all questions, make sure that when *formal inferential statistics are required*, each is preceded by the appropriate tests for the assumptions, i.e., state the assumptions, state the statistical procedure for testing the assumptions and mention their corresponding H_0 . If a graphical approach is used to test assumptions, explain the principle behind the approach. Explain the findings emerging from the test

of assumptions, and justify your selection of the appropriate inferential test (e.g. t -test, ANOVA, etc.) that you will use.

- **STATE HYPOTHESES** When inferential statistics are required, please provide the full H_0 and H_A , and conclude the analysis with a statement of which is accepted or rejected.
- **GRAPHICAL SUPPORT** All descriptive and inferential statistics must be supported by the appropriate figures of the results.
- **PRESENTATION OF ANSWERS** For each question, the answer must be written up in the format of a mini-paper under the section headings Introduction, Methods, Results, Discussion, and References. I don't expect that each answer would be more than 2-3 pages, although there are no lower or upper limits.
 - In this exam, you are required to provide any additional comments and supporting information related to methods, results, assumptions, and statistical explorations in a separate **Preamble** section, which will not be read as part of the mini-paper (i.e. it contains the 'behinds-the-scenes' workflows that do not make it into the paper). This section should contain any preliminary analyses, figures, tables, outputs, or any other relevant information not directly related to the formal hypothesis tests. **Please ensure to include the Preamble section prior to the Introduction section in your submission.**
 - The **Introduction** serves to provide background information, establish the context and relevance of the research, and clearly state the research question or hypothesis being investigated.
 - The **Methods** section will clearly outline only the statistical methods followed, e.g. which statistical tests were selected, how assumptions were tested, and a mention of any special data analyses that may have proceeded the statistical tests (if any). Typically, the focus here is only on the inferential statistics, not the EDA.
 - In the **Results** section you will focus only on the results around the hypotheses as stated in the Introduction. Although tests for assumptions also have statistical tests, they do not have to be mentioned in the Results.
 - The **Discussion** is where you will interpret and contextualise the findings, exploring their implications, limitations, and potential future directions within the broader scientific landscape around the topic. You can include up to five relevant papers across the Methods and Discussion sections.
 - A combined **References** section in the end after all the questions can contain all the references.

- **FORMATTING** Pay attention to formatting. Some marks will be allocated to the appearance of the script, including considerations of aspects of the tidiness of the file, the use of the appropriate headings, and adherence to code conventions (e.g. spacing etc.).
- **MARK ALLOCATION** Please see the [Introduction Page](#) for an explanation of the assessment approach that will be applied to these questions.

Submit the .html file wherein you provide answers to Questions 1–3 by no later than 08:00, Saturday, 13 April 2024. Label the script as follows:

BCB744_<Name>_<Surname>_Final_Integrative_Assessment.html, e.g.

BCB744_AJ_Smit_Final_Integrative_Assessment.html.

Email your answers to Zoë-Angelique Petersen by no later than 08:00 on 13 April 2024 and cc me in.

Question 1

Elephant growth data

Description Data on 288 African elephants that lived through droughts in the first two years of life.

Format A data set with 288 observations on the following 3 variables:

- **Age:** Age (in years)
- **Height:** Shoulder height (in cm)
- **Sex:** F, female; M, male

Background to the study can be found in the paper by [Lee et al \(2013\)](#).

The basic research question is whether there are sex-specific effects on growth of elephants.

The most basic answer is either “yes, there is a sex-specific effect” or “no, there is no sex-specific effect”. A substantive and statistically correct analysis addressing this most basic question will earn you 65%—note that a mark of 65% requires adhering to ALL requirements as per the instructions in the preamble. To get a mark of approaching 100% for this question will require additional analyses that demonstrate your own initiative towards achieving deeper insight into the biology of the species.

Question 2

For this question, you will obtain data on the effects of biochar on growth and elemental content of four crops, carrot, lettuce, soybean and sweetcorn from a US [EPA website](#). Please also consult the two papers cited there as some useful hints regarding the data analysis are available, which you might decide to heed (or not). You'll certainly want to read the papers for background to the studies.

The purpose of your work here is to focus on the plant yield and the three nutrients that you deem are most important in affecting human nutrition (the case for which must make in the Introduction section). Your analysis will allow you to make recommendations for about:

- i) whether or not there are differences between crops regarding the best biochar treatments to apply, and
- ii) to offer insight about how to best optimise the biochar application specifically for each crop with the aim to provide the best balance of human-benefitting nutrients produced and the biomass attained at the end of the growth period.

Question 3

Miscellaneous datasets are provided. Analyse each use the statistical test most appropriate for the respective datasets.

For this question, it will suffice to simply state the hypotheses in the Introduction and explain the findings and reach a conclusion in the Discussion. No need for referencing, contextualising, discussing, etc. The Methods and Results sections must be complete and detailed, however.

a. Spruce Moth Traps

Response: number of spruce moths found in trap after 48 hours

- Factor 1: Location of trap in tree (top branches, middle branches, lower branches, ground)
- Factor 2: Type of lure in trap (scent, sugar, chemical)

b. Apple Orchard Experiment

Five types of root-stock were used in an apple orchard grafting experiment. The following data represent the extension growth (cm) after four years.

- X1: extension growth for type I
- X2: extension growth for type II

- X3: extension growth for type III
- X4: extension growth for type IV
- X5: extension growth for type V

c. **Birds' Bones and Living Habits**

This dataset represent several ecological bird groups and measurments of various bones in their bodies.

Groups:

- SW: Swimming Birds
- W: Wading Birds
- T: Terrestrial Birds
- R: Raptors
- P: Scansorial Birds
- SO: Singing Birds

Measurements of bones (mm):

- Length and Diameter of Humerus
- Length and Diameter of Ulna
- Length and Diameter of Femur
- Length and Diameter of Tibiotarsus
- Length and Diameter of Tarsometatarsus

d. **The urine dataset**

This dataset is in the **boot** package and can be loaded as `boot::urine`. See the helpfile for an explanation of what's inside.