CSU : STA501
Statistics and Analysis
Feedback

Contents

Introduction

My name is Brendan Edwards (ID : 11759222) and I would like to provide feedback on the course notes of subject STA501 Statistics and Analysis. My hope is that the notes will be improved for future students and lecturers; and deliver a more positive experience for all going forward.

I undertook the Graduate Diploma of Applied Data Science in 2021 and 2022.

I found the notes in STA501 confusing. This not only made for unnecessary study overhead, made understanding the subject more difficult than it should have been, but also left a poor impression of the course and the university. I had suggested they be re-written, but the same notes were presented unchanged the second time. I figured this may have been seen as a disgruntled student complaining; so, I decided to record some of the issues I found.

The Study Guide, Interact Online Topic pages and Tutorials have

- Gaps,
- Inconsistencies
- Confusing notations
- Formatting that does not help understanding

Many of the issues I found are listed in the following pages. While doing the course, I ran out of time and interest to notate every issue due to the number of issues and frustration at having to notate them.

The beginning of every subject on the interact site, the outcomes state:

- *be able to examine critically and reflect on whether the statistical methodology and conclusions drawn in the media, scientific papers or reports are appropriate*

The way the course notes have been written and presented, there is an unnecessary overhead spent critically analysing the course notes, to then understand the content and the subject.

After discussing my grievances with friends and others that have also participated in the university sector, I realise there may be issues with preparation time and cost constraints.

My seat was subsidised as a this is promoted as a career transition course. If paying full fees of approximately $3,400, it would be even more disappointing and disheartening. Particularly if having to repeat to gain a qualification.

I feel it is important to provide positive feedback, but also express frustration at the quality of the notes, for a fee-paying course.

I would also like to suggest another resource for the course.

- ZStatistics by Justin Zeltzer, http://www.zstatistics.com/videos

Overall, I am happy and satisfied to have finally qualified for the Graduate Diploma. It certainly felt like an achievement to help further my career. The support staff and lecturers were all very helpful. I just want to make the course a more positive experience for a lot more people.

Please feel free to contact me if required.

Thank You and Regards,

Brendan Edwards

e) Brendan.edwards@understatedexcellence.com.au

m) 0407 526 745

Reference List

Outdated References (broken links)

Not all links are active.

- 2 out of 20. Usability and convenience.
- disappointing.
- Second year in a row

Topic 1

- Williams, N. (2014, October 8). Machine vision: The next wave of big data [Blog post]. Retrieved from https://www.flex.io/blog/machine-vision-next-wave-big-data

Topic 11

- Stephanie Glen. (2016, February 7). Bayes' Formula Example [Video file]. Retrieved from https://youtu.be/9miB7xbr59Y

Interact Online Lecture Notes

Download file naming conventions

- download file naming conventions
- inconsistent
- labelled weeks, instead of topics (due to mid-session break)
- can cause confusion when archiving and during revisions

Slide Pack titles

- download slide packs
- not all are numbered (by topic) or labelled (topic name) on the first slide
- 1, 2, 3, 4

Interact Learning Topics

Topic 1: Data and Microdata vs big data

Paragraph 3

Edit

FROM

--- --- ---

If data is collected on the characteristics of sampling units (i.e. observable objects) of a micro-population,

(such as a group of individuals, households or establishments) as collected by a census, survey or experiment,

then it is called microdata.

For instance, a national census might collect age, home address, educational level, employment status,

and many other variables, recorded separately for every person who responds.

This is microdata.

Typically, microdata is not readily available for analysis and/or modelling and needs to be simulated.

These sorts of data sets are beneficial for business organisations, policy makers and researchers.

A few examples follow:

--- --- ---

To – Suggestion

--- --- ---

A Census is a recording of a full population; similar to the Australian Census (https://www.abs.gov.au/census).

This can also be data from a survey or experiment; but it includes all of the data.

A Sample is a small part of the population and can be a subset that includes individuals, households or establishments)

Microdata are the characteristics from the Sample data.

Typically, microdata is not readily available for analysis and/or modelling and needs to be simulated.

These sorts of data sets are beneficial for business organisations, policy makers and researchers.

A few examples follow:

Big data wheel and big uncertainty

- no sub heading for Big Data Wheel
- there is one for Big certainty (usability)

Discussion Activity 3

- Students referred to Chapter 1 of the Study Guide: Exercise 3.
- no dataset provided or referenced.

Topic 3 - Basic Probability

Terminology and concept of probability

Edit

FROM

Calculating probability: If all the outcomes of the experiment are equally likely (i.e. 'all outcomes have the same chance of occurring'), then we say that the probability that an event occurs is equal to the proportion of outcomes in the sample space that belong in the event.

TO

Suggestion:

- If all OUTCOMES of an EXPERIMENT are equally likely to be observed
- When you specify a particular EVENT (combination of OUTCOMES)
- The PROPORTION of times that EVENT appears in the SAMPLE SPACE
- Is the same as the PROBABILITY

Example Questions from Textbook

- book provides answers for all odd numbered questions.
- reading material suggests EVEN questions.
- There are no answers in the text book to verify answers and understanding.

q6.24

- asks for a 'lurking variable' with no definition.
- this is an actual statistical term: Lurking Variables: Definition & Examples
- https://www.statology.org/lurking-variables/#:~:text=A%20lurking%20variable%20is%20a%20variable%20that%20is,relationship%20to%20appear%20to%20be%20present%20between%20variables.?msclkid=611c82d2b30d11ecbc0352a2da42b2d0
- Not listed in glossary provided

Topic 4

Browser tabs

- tutorial questions links to open in same browser tab
- tutorial solutions links to open in new browser tab
- other tutorials (so far) open in new tab

Useful Links

- has duplicated entries

Topic 5: Sampling Distributions

Section: Topic Overview

Paragraph 1

- Correct the html notation for mean and proportion =

Update text to include correct notation for web. It is displayed correctly in paragraph 2

- In this topic, we will look at the probability distributions of the sample mean \bar{X} and the sample proportion \hat{\rho}
- define 'proportion' and 'probability'. They can be read as interchangeable or notation for full population vs sample population.
- https://www.statology.org/probability-vs-proportion/?msclkid=3aa90819ad5f11ec8197423e05fc0099

Add diagram to reinforce written descriptions

- Sampling Distribution of the Sampling Mean
- add an explanatory diagram to reinforce the descriptive words

Section: sampling distribution of the sampling mean

Paragraph 1

- Simplify to notation only.
- Remove lots of excessive / unnecessary words

Quantitative

$n = sample\ size$

$\mu = mu = population\ mean$

$\sigma = sigma = mean\ of\ all\ sampple\ means$

Sample Proportions

$\bar{p} = population\ proportion$

$\hat{p} = sample\ proportion\ mean$

$n = sample\ size$

Topic 6: Inference based on a single sample

Estimation

Paragraph 1

- different symbol for definition of mean of sample compared to Study Guide

Paragraph 2
Edit
FROM

- *"...can be found in this topic's reading"*

TO

- *"...can be found in the Study Guide"*

Confirm correct referenced reading material

Hypothesis Testing

Paragraph 1

- update

FROM

- Section 5.3.2 of our reading discussed the case

TO

- Section 5.3.2 of the **Study Guide** discussed the case

Fully notate introduced symbols

Calculating confidence intervals for mu with non-normal data

- Alpha symbol α for confidence level introduced in equations but not defined in section
- other symbols previously defined and built upon
- no mention of confidence co-efficient as discussed in Study Guide, p.132

Topic 7: Inference based on two samples

Section: testing two independent samples when variances are equal

Paragraph 1

- Lots of words describing a specific test without naming the actual test

Name the test

"We consider such a test in section 6.3 of the study guide."

- *Study Guide:* ***6.3 Sample size for a desired margin of error***
- *Study Guide:* ***7.3 Testing 2 independent samples when variances are equal***

Confirm which is correct.

Also confirm which subsection 7.3.x is correct for correct referencing

- Comparison of Two means
- Comparison of Two proportions
- Comparison of Two variances
- Comparison of Two standard deviations

Comparison of population		2 sample t-test
Independent	No association	t-test : variance ≠ t-test : variance =
Dependent	Before v After	paired t-test

Topic 8. Analysis of Variance. Web Notes

Section: Analysis of variance (ANOVA)

Paragraph 2

- After table of data

"It seems fairly likely from the boxplots that feeding Regime 2 is "the best",..."

"The boxplots for the weight gains appears in Figure 7.1"

- There are no boxplots referenced or displayed to review and / or compare.
- Instead, the boxplots are listed in the Study Guide, Page 186, Figure 8.1: Weight gains of 20 piglets, five on each of four feeding regimes

ADD

- A reference to the location of the charts in the Study Guide, or
- include the chart in Interact at the table data

Section: The ANOVA

- List the variability types in the preferred (same) order

Compare to Topic 7 terminology

- Variability within (or inside) samples) – same sample - before and after
- Variability between samples (outside)

Section: Multiple Comparisons

Paragraph 2

- Use html for definition.
- Make rule and name more obvious
- Use vertical table to explain Type Errors instead of words

Section: How Large a sample do I need?

"For a single sample, we showed you in Chapter 5 of the Study Guide…"

- Repeat what is in Chapter 5 right here to give more context??

Section: Factorial experiments

Paragraph 4: "the most basic model…"

- Separate into dot points to make questions of interest stand out more

Section: Summary

- Describe how to get to each test in R Commander
- Ensure diagrams are consistent between the various chapters of the Study Guide and Interact Topic Notes

Topic 9. chi-squared test

Section: chi-squared test: independence of factors

Rewrite paragraphs to simplify

H_0

- no associations = independent
- actual cell counts = expected cell counts

H_1

- associations = dependent
- actual cell counts ≠ expected cell counts

Reject H_0 if statistics varies significantly (α) from expected value

H_1 is closer to the statistic than the expected value

- Paragraph 2 *"...supports H_1, then H_0 will be rejected."*
- Paragraph 3 *"The alternative hypothesis (H_a)..."*

While H_1 and H_a are effectively the same, and describe somewhere in the notes, select one and stay consistent throughout the course notes.

Section: chi-squared test: goodness of fit

Rewrite paragraphs to simplify

H_0

- expected frequencies agree well with observed frequencies

H_1

- expected frequencies do not agree well with observed frequencies

X^2 statistic increases as the difference between the observed and expected frequencies increase

Topic 10. linear regression and correlation

Section: Discuss

Update HTML list to be numbers instead of dt points to aid understanding

Section: multiple linear regression

"The technical contents of the multiple linear regression are described in the reading"

Suggestion.

ADD reference to actual description in specific reading material(s)

- *Study Guide chapter / page etc.*

Study Guide

Binomial Coefficient

- No final answers provided.

Calculations are there, but to make sure there is a full understanding of

- correct calculator usage
- correct manipulation of figures and numbers
- correct understanding of formulas and text...

There is no need for spoon feeding, but completeness helps understanding.

Poisson Distribution

- description of conditions = paragraph
- binomial description = dot points

Use dot points for simplification and ease of understanding

Chapter 4

page 98

Overview

The textbook adequately covers the required theory related to the Normal Distribution

Is the textbook the prescribed textbook, or the current Study Guide?? This can cause confusion.

Also mentions: knowledge of symmetry, looking up tables, basic arithmetic operations, addition and subtraction of areas are covered by year 9 secondary school...

ADD

- references to year 9 secondary books for a refresh or to gain an understanding.
- Or reference the prescribed text book at the appropriate pages and chapters.
- However, if you mention a year 9 book, then reference the appropriate book.

Using a shoelace example is a bit condescending and can make a potentially difficult concept make competent feel extremely dumb, incompetent and not capable.

GIVE

- references to fundamental learnings.
- There are people doing this course that are very capable that have not done detailed maths in a long time.

This course is a bridging course to a new career.

- Gaining an understanding of basic concepts to further career aspirations is incredibly important.
- There is no need to spoon feed; but there is no need to belittle people who want to learn.
- Year 9 text books are not part of the required reading, yet they are mentioned in this paragraph...

page 110

bottom of page. suggest rewrite

FROM

If you're waiting for something to happen, then - for times obeying the exponential distribution

- the fact that you have already been waiting s units of time makes it no more or less likely

that the event will occur in the next t units of times than if you had just started waiting now.

makes it no more or less likely that the event will occur in the next t units of times

TO

for times obeying the exponential distribution

if you are waiting for an event to happen

- Starting right now, OR
- After waiting for [s] units of time

It is no more or less likely that the event will occur in the next [t] units of times

Chapter 5: Sampling Distributions

Page 114

- second paragraph. Update /edit

Such quantities are functions of random variables, and so they themselves are random variables, because you cannot know with certainty what their values are until after the sample has been observed. Functions of random variables are called statistics. So, for example, X is a statistic, and so is S2. Please don't confuse `statistics' (more than one function of the sample values) with `Statistics' (the name of the subject). You can easily distinguish `pm' (post meridiem: after noon) from `PM' (Prime Minister') by the context, and it should be just as easy to distinguish `statistics' from `Statistics'.

Suggestion

- Simplify

Term	**First "S"**	**Definition**
Statistic	Uppercase	Formula
statistic	Lowercase	Parameter

Chapter 5

page 114

Rewrite or Remove

Please don't confuse `statistics' (more than one function of the sample values)

with `Statistics' (the name of the subject). You can easily distinguish `pm' (post meridiem: after noon) from `PM' (Prime Minister') by the context, and it should be just as easy to distinguish `statistics' from `Statistics'.

- disappointing, condescending, and irrelevant to the teaching experience.

NOTE

- if this form of explanation is needed, make the explanations much clearer and easily understandable at the beginning.

Chapter 6. Inference based on a single sample

Page 144, double negative

FROM

- 2. the variability in your data (the greater the variability, the less confident you can be that the sample is not misleading you).

TO

- 2. the variability in your data (the greater the variability, the less confident you can be in the sample data result).

Chapter 7

Summary, Page 182

page break should be before:

- Now we show a graph of the same information (including the chart from topic 5)

specify if topic 5 is the weekly csu interact topic.

NOTE. No similar chart listed in

- Study Guide Chapter 5

Interact topic 5. sampling distributions

However, a similar diagram listed in Interact topic 8. Analysis of variance

- chart, interact topic 8: does NOT have a line connecting 'n1 or n2 < 30' to 'unequal variances'
- diagram, study guide 7.6 summary, Paragraph 183: does have a line connecting 'n1 or n2 < 30' to 'unequal variances'
- diagram, study guide 8.5 summary, Paragraph 205: does NOT have a line connecting 'n1 or n2 < 30' to 'unequal variances'
- figure 8.8 specifies this is the flowchart for chapters 5, 6 and 7
- Are they the same chart as specified? Is the flowchart for testing meant to be different?
- If so, please specify the differences and summarise why. Thanks.

Page 187. Confirm correct reference

"Rough rule of thumb"

, largest sample variance is not more than four times the smallest sample variance, we will regard the population variances as being equal"

- *is not referenced in Chapter 6 of the Study Guide.*
- *referenced in Chapter 7, Paragraph 170*

Page 187, formatting. page break

the output of the R Commander is broken over two pages, with

- the headers on Page 187, and
- the data and description on Page 188

this can be easily remedied and improve comprehension by moving the page break to

BEFORE

- Paragraph 187. last paragraph. "Now let us consider the sample variances."

The Study Guide is delivered as a PDF document. That means the page limits have no restrictions. Use it to make the information clearer.

Readings and Resources

Update

- Stephanie Glen. (2016, February 7). Bayes' Formula Example [Video file]. Retrieved from https://youtu.be/9miB7xbr59Y
- Error Message: This video is not available anymore

Chapter 10: linear regression and correlation

Confirm correct referencing

- collinearity, Page 249
- we will discuss collinearity in subsection 9.5.5
- *there is no subsection 9.5.5*
- 9.x is on the previous section: chi-squared (x2) tests

Abbreviation of Terms / Acronyms

- 10.5.3. Hypothesis Testing, Page 246
- Lack of explanatory abbreviations
- Paragraph 1, "…simple linear regression model…"
- Paragraph 2, "In SLR we tested…"
- Paragraph 2, "However to test the multiple regression model equation…"
- Paragraph 2, "… the form of the Analysis of Variance (ANOVA) table…
- Paragraph 4, ".. unexplained by the MLR model…"

- At the start of the chapter, there is no abbreviation for the term Single Linear Regression (SLR).
- There is an abbreviation for the term: Multiple linear Regression (MLR)
- The abbreviation ANOVA for the term Analysis of Variance is used, even though this was described in chapter 8.
- There should be consistency of the use of terminology and abbreviations throughout the text. If ANOVA can be defined in Chapter 10, then so can
    - SLR and MLR, as can
    - SS_{xy} and SS_x and SS_y

Page 255, Method 2, naming of variables from table on Page 222

- Country
- Alcohol from wine in litres per person
- Heart disease death rate per 100 000 people

Described on ***Page 225***, Method 2 as

- Alcohol consumption
- Heart disease death rate

Suggestion

- make sure both labels and descriptions adequately describe what is being discussed.
- Or, use the same terminology in both instances.

Tutorials

- instructions for techniques are referenced in other tutorials.
- Recommend
    - including all relevant technique instructions in the actual tutorial.
    - these can/should/would be better to be all self-sustained resources - even as appendices.

Tutorial 4. Solutions

Question 5.

First sentence.

"Note. This question is based on random numbers and so your answer will NOT be the same as the example solution given here."

Unsure of the reasoning behind

1. NOT including the relevant material in a solution document to enable student to gain a proper and full(er) understanding of the material
2. Stating that the incorrect material has been provided
3. Causing confusion and disappointment in students undertaking the tutorial

Very Highly Recommended:

ONLY INCLUDE RELEVANT MATERIAL IN TUTORIAL SOLUTIONS

- Include the appropriate solution(s) directly related to question posed to aid in understanding of subject, topic and material.

Question 6.

First sentence.

"Note. This question is based on random numbers and so your answer will NOT be the same as the example solution given here."

Unsure of the reasoning or justification behind

1. NOT including the relevant material in a solution document to enable student to gain a proper and full(er) understanding of the material
2. Stating that the incorrect material has been provided
3. Causing confusion and disappointment in students undertaking the tutorial

Very Highly Recommended:

ONLY INCLUDE RELEVANT MATERIAL IN TUTORIAL SOLUTIONS

- Include the appropriate solution(s) directly related to question posed to aid in understanding of subject, topic and material.

TUTORIAL 6: inference techniques for two sample problems

Question 4 e

"To get a confidence interval, use the same commands as given in part ***Error! Reference source not found****"*

Tutorial 7: Solutions

Question 2

"In the supplied output, look at the Tukey confidence intervals…"

There is no supplied output included in the solutions document.

If the output is supplied in the Tutorial Questions document, how different documents are required to be open to complete the tutorials.

Why cannot the referenced output be included in the tutorial solutions?

Question 4 (two way anova)

Q) how to read f-tables to get DF=26

How is this done?

Turnit In : Plagarism Checker

- account expires even though I am registered for a full session.
- I had to register for an account to check and submit assignment 1
- my account had expired when I went to submit assignment 2
- the only option I had when logged in was for the 2021 classes.
- incredibly annoying and not mentioned in notes.
- I would have expected that being logged in for a 6-month class would not have the account expire every month or so.
- apparently, my account expired on February 26, even after I re-registered for the 2022 class.
- ~~there is no upload button described.~~