

Statistical Modelling Assignment

1 OVERVIEW OF THE ASSIGNMENT

This assignment will test your skill to collect and analyse data to answer a specific business problem. It will also test your understanding and skill to use statistical methods to make inferences about business data and solve business problems, including constructing hypotheses, test them and interpret the findings.

In Australia, many people need to lodge a tax return after the end of the financial year. They can prepare and lodge their own return or pay a registered tax agent to do it for them. By using a subset of the sample file from the Australian Taxation Office (ATO), your task is to summarise and analyse several aspects of this lodgement method. We are interested to know the proportion of people who lodge a tax return using a tax agent; whether there is a difference among the age groups in terms of their lodgement method; whether there is a relationship between total income and the lodgement method; and the relationship between total income and deduction amount. In addition, you are also asked to collect and analyse a dataset about international students' preference of tax return lodgement method.

2 TASK DESCRIPTION: WRITTEN REPORT

There are two datasets involved in this assignment: Dataset 1 and Dataset 2, detailed below.

Dataset 1: You will receive an email that contains a dataset that is specifically allocated to you. This dataset is a subset of 2013-2014 individual sample file, provided by the ATO and has been edited to only include a subset of the cases and variables. The original dataset can be obtained from <https://data.gov.au/dataset/taxation-statistics-individual-sample-files>, and it is under the license of [Creative Commons Attribution 3.0 Australia](https://creativecommons.org/licenses/by/3.0/au/). Data dictionary of the edited dataset is given in the following table.

Variable name	Description	Values
Gender	Gender (sex)	0 = males, 1 = females
age_range	Age in five years ranges	0 = 70 and over 1 = 65 to 69 2 = 60 to 64 3 = 55 to 59 4 = 50 to 54 5 = 45 to 49 6 = 40 to 44 7 = 35 to 39 8 = 30 to 34 9 = 25 to 29 10 = 20 to 24 11 = under 20
Lodgment_method	Lodgment method	A = Tax Agent

		S = Self Preparer
Tot_inc_amt	Total income	All numeric
Tot_ded_amt	Total deductions	All numeric

Dataset 2: Collect data (e.g. via a survey) from international students about whether they would use a tax agent to lodge a tax return in the future. There is no requirement about sampling methods and sample size, but you need to justify your approaches in Section 1 (see below).

Both datasets should be saved in an Excel file (one file, separate worksheets). All data processing should be performed primarily in Excel or by using Statkey tool.

Prepare a report in a document file (.doc or .docx) which includes all relevant tables and figures, using the following structure:

1. Section 1: Introduction
 - a. Give a brief introduction about the assignment, and include a short summary of a related article with a proper citation.
 - b. Dataset 1: Give a short description about this dataset. Is this primary or secondary data? What types of variable(s) is involved? Display the first 5 cases of your dataset.
 - c. Dataset 2: Explain how you collect the data and discuss whether your sample is biased. Is this primary or secondary data? What type of variable(s) is/are involved? You don't need to display your data in this section.
2. Section 2: Lodgement Method –
Dataset 1 Use Dataset 1
 - a. Using suitable graphical displays, describe the variable lodgement method for Dataset 1.
 - b. Calculate a 95% confidence interval of the proportion of tax payers who lodge the tax return by using an Agent.
 - c. Give a short comment about your finding.
3. Section 3: Lodgement Method –
Dataset 2 Use Dataset 2
 - a. Using suitable graphical displays, describe the variable lodgement method for Dataset 2.
 - b. Calculate a 95% confidence interval of the proportion of tax payers who lodge the tax return by using an Agent.
 - c. Compare this result with the result in Section 2 and make a comment whether there is a difference between dataset 1 and dataset 2 in terms of lodgement method.
4. Section 4: Lodgement Method and Age
Group Use Dataset 1
 - a. Describe the relationship between the age group and lodgement method using suitable graphical display and numerical summary.
 - b. Perform a suitable hypothesis test at a 5% level of significance to test whether the two variables are associated.
 - c. Give a short comment about your finding.
5. Section 5: Lodgement Method and Total Income
Amount Use Dataset 1
 - a. Describe the relationship between total income and lodgement method using suitable graphical display and numerical summary.

- b. Provide a comment about your result in part a (include a comment about the shape of the distribution, centre, spread and outliers).
- 6. Section 6: Total Income Amount and Deduction Amount
 - a. Describe the relationship between total income and total deduction using suitable graphical display and numerical summary, for each type of lodgement method.
 - b. Provide a comment about your result in part a.
- 7. Section 7: Conclusion
 - a. What can you conclude from your findings in the previous sections?
 - b. Give a suggestion for further research

3 TASK DESCRIPTION: PRESENTATION/INTERVIEW

A presentation/interview for the assignment is scheduled on Week 11, in your allocated tutorial.

You do NOT need to prepare a presentation material (e.g. power-point slides), instead, you will be asked to demonstrate and/or explain how you summarised the data and how you performed the analysis. You may be asked to replicate what you have made in your written report (e.g. generate a chart or numerical summary using Excel or Statkey).

4 SUBMISSION REQUIREMENT

Deadline to submit written report: Week 10 Wednesday (24 January 2018), 5pm

You need to submit 2 files to Turnitin:

1. Main report, in a Microsoft Word document file
2. Dataset, in a Microsoft Excel file

Main report (word document):

1. Size: A4
2. Use cover page with your details and signature
3. Single space
4. Font: Calibri, 11pt

Dataset (excel document):

1. Dataset 1 in Sheet 1
2. Dataset 2 in Sheet 2
3. Pivot tables and any other information in other sheets (rename the sheet appropriately)

5 DEDUCTION, LATE SUBMISSION AND EXTENSION

There is a 2-mark deduction (out of 20) for students who do not address the specification in the submission requirement listed in Section 3 above.

Late submission penalty: - 5% of the total available marks per calendar day unless an extension is approved.

For extension application procedure, please refer to Section 3.2.1 of the Subject Outline.

6 PLAGIARISM

Please read Section 3.3 Plagiarism and Referencing, from the Subject Outline.

Below is part of the statement:

“Students plagiarising run the risk of severe penalties ranging from a reduction through to 0 marks for a first offence for a single assessment task, to exclusion from KOI in the most serious repeat cases. Exclusion has serious visa implications.”

“Authorship is also an issue under Plagiarism – KOI expects students to submit their own original work in both assessment and exams, or the original work of their group in the case of a group project. All students agree to a statement of authorship when submitting assessments online via Moodle, stating that the work submitted is their own original work.

The following are examples of academic misconduct and can attract severe penalties:

Handing in work created by someone else (without acknowledgement), whether copied from another student, written by someone else, or from any published or electronic source, is fraud, and falls under the general Plagiarism guidelines.

Students who willingly allow another student to copy their work in any assessment may be considered to assisting in copying/cheating, and similar penalties may be applied. ”