The University Of Sheffield.

INF6029 Data analysis

Advice and suggestions on individual report

Based on past common mistakes!

The University Of Sheffield.

General points

- Avoid the phrase “the elderly”, it is ageist
  - Use ‘older people’ or ‘elderly people’
- You should not imply causation from cross-sectional analyses. Therefore:
  - Avoid these words:

    “impact on”; “cause”…
  - Use these instead:

    “associated with”, “risk factor for”…

The University Of Sheffield.

General points

Avoid the following:

- Colloquialisms:
  - e.g., "ballooning numbers of people who feel lonely"
- Americanisms:
  - e.g., use children not ~~kids~~, colour not ~~color~~, analyse not ~~analyze~~
- Contractions:
  - E.g., use "do not" not "~~don't~~", "cannot" not '~~can't~~', etc.

The University Of Sheffield.

General points

- Write body of report in the past tense (because when you submit you have already carried out the research!)
- Add numbering and captions to all figures and tables - bold, above, left aligned
- Refer to figures and tables within the text
- Avoid floating sentences between tables (chi-square especially). Aim to write in paragraphs/sections
- Check for proofreading errors - e.g., missing spaces, repeated sentences, wrong capitalisations, etc.
- Make sure you cite all resources in the correct APA referencing style (https://librarydevelopment.group.shef.ac.uk/Assets/pdfs/referencing/apa-7ed.pdf)

The University Of Sheffield.

General points

- Students sometimes have difficulties interpreting variables that were significant using chi-square but not significant in combination with others (logistic regression). Often they implied that the logistic regression results meant the chi-square test results were invalid. However, this is not the case, but the discrepancy could be because all tests have specific rules that you need to fulfil: sampling distribution, sampling size, type of variable, etc. The interpretation depends on the analysis and different results can happen!
- Avoid choosing variables that are not appropriate for your research question. E.g., "whether or not they did some exercise yesterday" cannot be used to imply the general "exercise level". Or: "whether participants retired early" cannot be used to imply "early retirement“. In fact, some participants were not even at retirement age and therefore might have answered "no" at this point simply because they weren’t retired (yet).
- Use your imagination when setting up your research question! There are not just depression, happiness, and life satisfaction!
- What to do if things have gone wrong? Often students would 'admit' a fundamental flaw in their analyses at the end of the report - better to pull apart why this invalidates the results earlier to show understanding.

The University Of Sheffield.

Introduction, lit rev, research aim and objectives

- The **Introduction** needs to fully justify the selection of a specific topic, provide context and background information
- The **Literature Review** should be a critical appraisal of the literature in which individual topics are intertwined to tell a story - not like a “shopping list” - often commented to ask 'why' rather than just 'what‘
- Both Introduction and Lit Rev cannot have unattributed statements (i.e., with no references for evidence/support)
- Use current (possibly from the last 5 years), peer-reviewed literature
- The **aim** needs to express in one sentence what the main purpose of the report is
- The **objectives** must not be methodological steps. Instead, they break down the general aim into smaller, logically connected parts that systematically address the various aspects of the overall problem

The University Of Sheffield.

Results

- DO NOT copy and paste the output from SPSS tables into the main report – instead, compile results in Word or Excel tables
- Descriptive statistics:
  - Include age (groups) and gender as a minimum for demographic variables
  - Make sure you include the age group ‘90+’ (i.e. consider merging “Age” into age groups: e.g. 50-59, 60-69, 70-79, 80-89, 90+ or other groupings)
  - Include *n* values in tables and text (as well as % values)
  - Check the numbers in the descriptive statistics - are there a lot of missing data?
  - Interpret missing data - it is not enough to just report them in tables
  - No need to recode already categorical variables - there are other ways of removing missing data
  - If you need to re-code a variable, mention how you did the coding (e.g., 7-point Likert Scale can be coded into two groups but you need to say what is included in Group 1 and Group 2)
  - When reporting descriptive statistics in-text try to 'always add something new‘ instead of repeating the results in the tables

The University Of Sheffield.

Results

- Inferential statistics:

In Chi-square tests:

- Report ALL outputs (including crosstabulations!!) in the report
- Assess the validity assumption based on the expected cell counts
- Use ‘Continuity Correction’ for 2x2 tables
- Use ‘Linear-by-linear association’ for ordinal variables
- Use ‘Pearson Chi’ for nominal variables
- Do not report ALL test statistic values in the text, but only the one relevant to the specific crosstabulation you are discussing
- Include cross-tabulations with row %s **or** column %s (not both!)
- Report the direction of any associations (the relative proportions) in different categories
- Provide the interpretation of Odds Ratios that are <1 (i.e., not just OR>1)

The University Of Sheffield.

Results

- Reporting binary logistic regression results:
  - Report the Nagelkerke R^2 value
  - Provide the category names (e.g., Excellent, Good, etc.). Not just the dummy variable labels (1.0, 2.0, 3.0, etc.) as this does not mean anything to the marker
  - Mention the reference category for each variable and what you are predicting (e.g., are you predicting "no depression" or "depression“?)
  - Explain the increased (OR>1.00) or decreased risk (OR<1.00) associated with independent variables/categories
  - Report the output tables in the correct order (overall model statistics, dummy variables and regression model)

The University Of Sheffield.

Discussion

- This should not be another ‘results’ section. This section involves assessing your general results in relation to past literature - no need to repeat the actual numerical outputs
- Fully engage with the literature - similarly to Introduction and Literature Review sections, the Discussion needs to be fully supported by the literature
- Remember to add meaningful limitations and suggestions for further research