Text Analytics for Surveys
PASW
A Brief Introduction

Dr Helen Klieve
Stats Advisor
School of Education Brisbane/Logan
h.klieve@griffith.edu.au
Presentation Summary

- Using surveys in educational research
- Accessing text information
- The TAS Process
- Learning's
- Live example
Acknowledgments

License provided by Malcolm Wolski

Director, Information Services, to trial the system for use in Griffith.

This is the first communication of the approach.

Also to the Survey Research Centre staff for support in doing on-line surveys.
Survey Analysis:

- Surveys capture a range of data including:
  - Demographics/Explanatory vars. Numeric (N)
  - Opinions/Actions – Questions and Scales (N)
  - Measures – Scores, Test results (N)
  - Clarifying information (Text (T))
  - Views from open questions (T)
The Quantitative Qualitative divide

Is it

Quantitative Research

OR?

Qualitative Research
The Quantitative Qualitative divide

Research Question addressed by

Quantitative Tools

Qualitative Tools

>> Evidence making a case
Example Data Set:

Evaluation of Training
Objectives in participation /Other Issues

- Demographics
- Community Issues Profile
- Leadership Scale

**Contribution of Qualitative:** Open Response field

- Interest between 3 “theme areas”
  - Leadership
  - Networking
  - Mental Health
- Operational issues – in the training
- Other broader issues
Often as quotes/summary points, eg:

Mental Health

- Assistance with learning to look after my own mental health and the mental health of others in my community and network
- To learn about signs of mental health problems and help those people out there so not alone.
- Awareness of mental health and how to deal with it in my community.
- To be understanding of what to do and not to do in situations
- Develop skills to help myself and friends overcome mental health issues.
- Ability to be helpful and recognise poor mental well-being, depression. And give some direction for professional and non-professional support. Looking forward to mental health first aid

The above comments highlight the consistency of interests with the overall stated objectives of the program. Some of the feedback provided reinforces the success of the program for the participants.

Participant assessments

- Thanks for coming to a town of ..... people!
- Well done. Training is good to grow yourself continuously
- A very necessary learning experience for this area. I thoroughly enjoyed it. Helped me to understand parts of my past
- Thanks - this has been a great course - very helpful
- Please advertise and present this course again and I would like to say that this program has helped me enormously and I will be advising others to make use of a similar program should it happen again
- I enjoyed the 2 sessions I was able to attend - learnt a lot
- Greater understanding of difficulties facing rural and remote people
- I think this program was wonderful. It gave me so much to think about and new ways of working in my community
- Excellent speakers and educators, easy to understand- have learnt a lot of new information

This does provide some voice of the participants but more effective use of this form of information could be made
Surveys – on-line responses

- Increasingly important means of collecting views and opinions
- Have the capacity to capture responses to closed questions but also open questions and “open responses”
- Thus in LimeSurvey have:
  - Short Free Text
  - Long Free Text
  - Hugh Free Text
What sort of responses can you expect?

- “Electronic survey respondents are more likely to be self-absorbed and uninhibited when they complete a survey by computer – and may concentrate more on the questionnaire”  
  (Kiesler and Sproull 1986)

- “A number of researchers ...have reported that respondents write lengthier and more self-disclosing comments on e-mail questionnaires than they do on mail surveys”  
  (Yun & Trumo, 2000)
Surveys – using on-line approaches

- Confidentiality – this supports accessing real views
- Many people are happy to provide detailed and honest views
- You can access a far greater range of views than through 1:1 interviews with much larger samples

How to Analyse/How to Report
Examples of level of response – Autism Survey:
74 of 91 respondents (81%) provided a detailed comment on 1/both the options
currently much of the 'transition' training for our students is from prep to year 1 and the prep classrooms are largely responsible for the environmental transitioning of our students. Prior to prep our focus is on skilling students in self care (toiletting, food times, possession management), developing language (reciprocal conversations, question formats, engaging peers, conflict resolution), attending to task and maintaining this attention in short cultivating classroom etiquette. But we haven't formalised the process to – can these things be better? problematic - we are directed to not 'place' students, it is to be parent driven except for early entry special school where the process is very clear and straightforward. Schools feel 'not ready' for our students until they arrive the following year. Information and files are often misplaced and staffing isn't finalised and there is no clear, quick, effective and consistent support to prep students. schools vary greatly as to how this is handled. BUT this questionnaire has given me much food for thought with regard to some of our practices and areas that need attention. thank you
Can we better target & use this information?

Open questions are powerful:
- They can avoid prompts – the first opinion
- This can also identify themes where these aren’t clear (eg grounded theory approach)

Can also be interesting to see not just the range of responses, but who makes which comments

*Are there better ways to plan, ask and analyse this data?*
How can we approach the analysis?

- Do formal analysis – eg grounded theory design? (large sample/limited individual detail)
- Look at responses and identify useful examples of “participant voice”;
- Read each response and code eg Nvivo
- Use language analytic techniques
  - Eg Leximancer – focus at the document level
  - Text Analytics – focus at the respondent level

But no magic
What might TAS analysis look like?
A summary of the frequency of issues
Presentation Options

Circular Layout (each respondent counted once)
Do people in different communities perceive “Mental Health” with the same priority?

Interest in Mental Health
% identifying

<table>
<thead>
<tr>
<th>Location</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mt Isa</td>
<td>25</td>
</tr>
<tr>
<td>Longreach</td>
<td>40</td>
</tr>
<tr>
<td>Kingaroy</td>
<td>45</td>
</tr>
<tr>
<td>Roma</td>
<td>5</td>
</tr>
</tbody>
</table>
The TAS Process
Text Analytics for Surveys

1. Collect data – enter into SPSS/PASW
2. Open TAS environment
3. Start Project – identify relevant fields (ID/T)
4. Decide how to “Build” – concepts/patterns
5. Build categories:
   1. Re-define & Re-group; or
   2. Force
6. Do analysis/graphics
7. Export – as “binary”/ link to other SPSS tools
8. Extended analysis in SPSS
Go into TA

- Select NEW
- Identify SPSS file (or excel)
- You MUST have a unique ID (it won’t accept duplicates)
- There are constraints with very large data storage requests
- DO NOT delete SPSS file – it stays linked!!
Move ID

And the text variable/s

Into the boxes to define the analysis
The initial extraction

- Pre build
- All responses
- All empty categories
- Extractions by concept
The initial “BUILD”

“Built” categories

Responses that load on a built category
After initial **BUILD**

Groupings after Build

Responses with links to identified group

Original grouping
Initial graphic - circular
Mapping of themes and interactions
Are these the right categories?

- Membership
- Wording
- Completeness?
Recoded

- Using initial themes
- “Force” categories – link to relevant responses
- Check each response for membership
- Observe Frequency Histogram
- Observe “Layout” options, showing main themes, and also interrelationships. Note each participant contributes only 1 weight.
A summary of the frequency of issues
Presentation Options
Circular Layout
Directed Layout
Grid Layout
Data can then be moved back into an SPSS environment (or excel)

- File
  - Export
    - To PASW Environment
New SPSS file with codes by ID in columns (1 if referred)
Do people in different communities perceive “Mental Health” with the same priority?

Interest in Mental Health

% identifying

Mt Isa
Longreach
Kingaroy
Roma
Leadership vs Interest in Mental health

- High: No > Yes
- Moderate: No > Yes
- Low: Yes = No

Concern with Medical services vs interest in mental health

- No: v v major concern, v major concern, major concern, concern, neutral, strength, major strength, v major strength, v v major strength
- Yes: v v major concern, v major concern, major concern, concern, neutral, strength, major strength, v major strength, v v major strength
Learnings

- Major benefits over Nvivo/Leximancer:
  - Analysis is at the response level;
  - You can look at response patterns; and
  - You can link response patterns to demographics.

- It's easy to get a very simplistic result – you need to “get inside” your data to really make the most of it

- With more clearly scoped open questions it will be easier to apply automatic coding
Let's look at operational examples