

# *Automated vs. manual methods of coding and analysing free text survey responses*

Dr Kathy Seymour, Director,  
Seymour Research Ltd



# Free text data

General

- Please use the space below to provide any further comments on the topics covered in this survey

- Please describe below the most positive aspect of your experiences at this university

- Please use the space below to suggest any improvements that could be made to your course

- Please use the space below to comment on overall teaching standards on your course

Specific

- Please use the space below to describe what you liked best about the teaching methods used on module ENG3001

- We are looking at free text data generated in response to open-ended survey questions
- These can be general or very specific as shown in the examples on the left
- The degree of generality is important – we'll come back to this later!

# Data used in this research

- For this research we used comments from undergraduate students who completed the 2012 National Student Survey
- Students are asked:  
*Looking back on the experience, are there any particularly positive or negative aspects you would like to highlight?*
- We took only the **negative** comments for this research
- In total there were **1329** comments in this dataset



# Text analysis tools used



“NVivo is software that supports qualitative and mixed methods research. It’s designed to help you organise, analyse and find insights in unstructured, or qualitative data like: interviews, open-ended survey responses, articles, social media and web content.”



“IBM SPSS Text Analytics for Surveys (v4.0.1) lets you transform unstructured survey text into quantitative data and gain insight using sentiment analysis. [It] uses natural language processing (NLP) technologies specifically designed for survey text. [It] categorises responses and integrates results with other survey data for better insight and statistical analysis.”

# Results from NVivo

Eight overall themes were identified (parent nodes) and within each several sub-topics (or child nodes) were also identified:

- Course
- Quality of teaching
- Facilities, resources and support
- Assessment and Feedback
- Placements or work experience
- Careers advice and employability
- General positive/negative comments
- Online learning



Topics within 'quality of teaching'	n
Staff support and contact not good enough	107
Lecturers' teaching abilities need improving	57
Generally negative comments about the quality of teaching	32
Dissertation support has been poor	26
Lecturers' communication and/or language skills are poor	24
Lecturers' knowledge and/or expertise needs improving	16
Personal tutors have been poor	9
Variable teaching quality	8

# Results from SPSS Text Analytics

- Using the 'concept root derivation' categorisation method STAS identified 40 level 1 codes
- The table shows those mentioned by more than 30 different respondents

Higher level theme (category)	No. of respondents who mentioned this
class	173
students	71
modules	65
time	62
course	56
tutoring	52
lectures	50
communication	46
feedback	43

Higher level theme (category)	No. of respondents who mentioned this
placements	42
teaching	40
tutors	39
lecturers	39
staffing	38
assignments	36
work	36
computers	34
timetable	31

# Results from SPSS Text Analytics

Some of these level 1 codes appear to refer to quality of teaching:

- Tutoring (n=52)
- Lectures (n=50)
- Teaching (n=40)
- Tutors (n=39)
- Lecturers (n=39)

Also possibly some aspects of

- Class (n=173)
- Staffing (n=38)

Total 431

Compared to

279 in NVivo

And we don't know *what* was said about each topic

# Results from SPSS Text Analytics

Looking at comments categorised under 'teaching', many of these did refer to poor quality teaching

*Some of the lecturers were actually very bad at teaching and the whole class found it a struggle sometimes*

But others did not directly refer to teaching quality, instead these were issues that would have been best highlighted elsewhere

*Not enough teaching which is what you pay for*

This would have been better coded under a 'value for money' category



# Comparing the methods



- ✓ Good for a thorough analysis of the data, especially with inter-coder reliability checks
- ✓ Good for long comments
- ✓ Lots of additional analysis techniques / functions
- ✓ The 'human element' reduces error and ambiguity
- ✗ Time consuming
- ✗ Less suitable for very large datasets



SPSS Text Analytics for Surveys

- ✓ Good for large datasets
- ✓ Good for shorter comments and/or very specific survey questions
- ✓ Good if you do lots of similar surveys
- ✓ Good to analyse against tick box variables
- ✓ Quicker than 'manual' coding
- ✗ Large amount of user 'fine tuning' required and this can have a dramatic effect on the resulting analysis
- ✗ Steep learning curve

# Conclusion

- I would usually favour NVivo and there is additional functionality in NVivo 11
- If large dataset / lack of resources is a problem, consider analysing a sample in NVivo
- The human element means analysis can focus on the research question(s)
- STAS is fine for short text responses, or if you are trawling lots of data for a specific topic, e.g. timetabling
- STAS is expensive and appears to be a low priority for IBM
- It's worrying that the two techniques elicited such different results
- Automated methods are not necessarily a shortcut – lots of user intervention required and this often requires a good knowledge of your data, so you end up chasing your tail



# Thank you



*Questions and discussion*