A Guerilla Experiment In Sentiment Analysis

Syncapse Measurement Science Team

Christopher Berry, Kevin Richard, Hemash Bhatti, Gregory Araujo, Zoe Siskos and Nadia Kim

A Guerilla Experiment in Sentiment Analysis

Positive, Neutral and Negative

Abstract

The summarization of large amounts of information into a single digestible Key Performance Indicator is a major problem in social media measurement. The dominant paradigm in social listening tools is to rate all content with a positive, neutral, or negative value. We call this the Generic Sentiment Score. A major source of dissatisfaction among marketers and PR professionals is the overall accuracy generated by this score and the machines that try to approximate this value. This experiment proves that there are many versions of truth among humans when asked to score the same content. The authors conclude that the Generic Sentiment Score is not enough and that a more comprehensive approach is required.

Introduction

A major problem in social media measurement is that of sentiment and tone analysis in User Generated Content (UGC).

People are producing more information than ever and sharing it online through social media. People share, tweet, update, comment, post, blog, forum, troll, photoshop, upload, download, and rate. It is a lot. And marketers, like all people, seek a simple way to sum it all up.

Marketers rely on seemingly simple metrics to judge performance and make decisions. TV has Nielsen Points. Print has circulation. Web Analytics has conversion. Social Media has...the Generic Sentiment Score.

We have observed that most social listening tools produce a Generic Sentiment Score based on different algorithms, but in general, they all sum up a piece of content as being positive, neutral, or negative. Most of these algorithms are based on keyword matrices – where the appearance of the word 'hate' might be worth -8 and 'love' might be worth +8. When all the words in a piece of content are summed up and then averaged, one gets the Generic Sentiment Score. Average the Generic Sentiment Score across all the blogs, posts, tweets, and comments – weighed or not – and one is presented with the Generic Brand Sentiment Score. Even though this averaging conceals variability among the statements, we are told that positive is good and negative is bad and advised to take action to make overall sentiment improve. There are variations in how some sentiment engines score words. Some rely on a human trainer. Others rely on keyword frequency density. Some rely on neuroanalytics. Others are entirely human scored. All machine algorithms, and as we will demonstrate, human judgment will generate error. This error can be crippling when the magnitude and direction isn't fully understood.

The dissatisfaction with any given score is driven, in part, by personal subjective experience – typically by reading a piece of content in a social listening tool and comparing it against what the machine says it is. The typical reaction is "that's not right"! This dissonance between instrumentation error and personal subjective perception becomes especially acute when the machine is not painting a pretty picture.

The quality of machine algorithms is not solely to blame.

If you were to ask three people how they are feeling today, you could potentially get three different responses. The first being that the person is feeling 'okay', the next person says that they are 'better than yesterday', and the third person answers by telling everyone that they are doing 'great'. At first glance, all three statements can be considered as positive statements. However, when you analyze the answers further, 'okay' can mean that the person is neutral, 'better than yesterday' can mean that yesterday was a horrible day and today is not as bad, and the third response can mean exactly what they say[1]. If humans themselves cannot unanimously agree on the tone of these three statements, then what chance does a machine have?

As this experiment will demonstrate, Generic Sentiment Score should not be taken as a definitive conclusion. In turn, sentiment score should not be the only indicator when making actionable recommendations.

The experiment began on January 20th, 2010, for a five-day period where participants were invited by way of a Tweet. The experiment obtained 112 respondents. There were 10 incomplete questionnaires, which were not included in the final analysis.

We call it a guerilla experiment because of the short time period it was executed in and in a popular spirit. It was a piece of self-directed research that the Syncapse Measurement team wanted to take on.

The team wrote five hypotheses to investigate Generic Sentiment. The first two hypotheses proved that no single question received unanimous agreement of a sentiment score. Moreover, it was found that no two questionnaires were scored exactly the same by 102 humans. In addition to investigating the sentiment scores, we also evaluated other potential factors that would impact the overall rating. These factors included the mood of the scorer, the "anchor and adjust effect" and the influence of rating the first five texts as negative. Other factors exist and we chose brevity of the survey over isolating more of them – such as the differences among people and within people. We found only a weak relationship existed with these variables on the overall sentiment score. Although not a strong influence, the anchor and adjust effect had somewhat of an impact if the responder rated the statement as positive, neutral or negative.

A Background on Sentiment Analysis

The sentiment analysis problem in social media is not a unique or original one. The quantification of text information, in particular, bias, is common in the fields of linguistics, neuroscience, developmental and industrial psychology, content analysis, communication studies, marketing science, and certain branches of political science - among many others.

We re-read the original paper by Tversky and Kahneman (1974) who examined the different mental heuristics people use in decision-making. One mental heuristic that the authors focus on is that of anchoring and adjusting. They did a study to examine the effect of anchoring on human perception.

They wanted to know if just by hearing a number, regardless if it was related to the topic at hand, would influence the a persons answer to another question. It would not matter if the one figure had anything to do would the next. The experimenter spun a wheel to derive a random number to give them a starting value. Subjects were asked a variety of questions (for instance, the percentage of African countries represented in the United Nations). Each time, subjects' estimates were influenced by the initial value provided by the random spinner. If the random number was high, people estimated that a large percentage of countries were represented in the UN. If the number was low, their estimates tended to be lower. In effect, people's perceptions were anchored on a figure, from which they had to adjust.

[1] The variation of word meanings among people compounds the problem. The word 'sick' means something to a Banff 26 year old snowboarder and a Toronto 57 year old information worker. In certain contexts they would not agree on the meaning.

We hypothesized that the same anchor and adjust impact would be reflected in the answers over the course of a survey. We had observed the impacts of 'push polling' before, and wondered if somebody perceived something as negative would influence their response to the next question. The impact is that machines are not guilty of anchor and adjustment errors – but humans could be.

The second is Carrabis' book, "Reading Virtual Minds Volume 1" (2009). In it, Carrabis details how language is reflective of cognition, and through language, one can not only understand what somebody is thinking, but also predict the likely impact on other people[2].

The third author we want to cite is a sequence of opinions stemming in social media measurement relating to negative, neutral, and positive sentiment. Francois (2009) summarizes the opinions of French researcher Guilhem. Guilhem proposes that sentiments contain gradients not measured by current tools including social connectivity. Also, Gattiker (2009) states the weak validity present in current sentiment analysis practices, as they do not factor in linguistic differences among cultures such as those between the United States and the United Kingdom.

We want to demonstrate that a panel of humans, given equivalent ambiguous instructions that most marketers are given when they log into a listening platform, could not unanimously agree on the Generic Sentiment Score[3].

Methodology

A survey, hosted by Google™ Forms, was deployed between January 20th and 25th 2010. Although 112 questionnaires submitted of which 102 were completed from Q1 to Q25. The analysis only includes the data of the completed questionnaires.

Human respondents were invited to participate by way of a Tweet. It was subsequently re-tweeted. An email was sent out by way of Syncapse inviting people to participate, and inviting anybody else who might be interested to participate.

The questionnaire (Referenced in Appendix 9) opened with the following text:

"Thanks for clicking on the link! The following experiment should take 5 to 8 minutes.

We want to share both the results and the dataset from this study openly.

At the end of this experiment, the results of this survey – the full CSV datafile - and accompanying paper will be published on syncapse.com, christopherberry.ca, and kevrichard.com. The authors are not keeping a record of anything personally identifiable. They are not asking for your name or email address.

Please indicate whether you believe the statements below are positive, neutral, or negative - towards books."

The questionnaire comprised of 25 sample tweets related to the topic of books. Followed by this was a question related to mood, and then an open text box so responders can provide other comments on how the text was scored.

[2] The effects of opinions on people, and the subsequent word of mouth that may follow, has a lot to do with the speaker, listener and context. A quantitative understanding of those effects in part relies on the derivation of relatively stable text metrics and is worthy of future study.

[3] It's possible, though improbable, that enough of the variances among humans could be isolated such that a consistent, valid, and 100% accurate sentiment algorithm could be derived.

To mathematically compute the Generic Sentiment Score, we assigned all answers negative a value of -1, neutral a value of 0, and positive a value of +1. We added up all the scores to derive the Generic Sentiment Score – which would range from -25 and +25. A score of -25 would mean that a respondent had scored them all negatively, and a score of +25 would mean that the respondent had scored them all positively.

We decided not to average it out so as to keep it as simple as possible to understand.

Experiment Overview

The purpose of this experiment was to examine Generic Sentiment Scores.

Five hypotheses were investigated:

Hypothesis 1: There is no unanimous agreement among all respondents on the sentiment score of a single response to any question within the survey.

Hypothesis 2: No two completed surveys will contain identical responses. Every completed survey will be unique.

Hypothesis 3: The reported mood of the scorer and the Generic Sentiment Score are correlated such that negative moods generate negative scores.

Hypothesis 4: If the responder starts off rating the first five statements as negative, the subsequent responses will tend to be more negative than those who started off positive.

Hypothesis 5: The tendency to anchor and adjust causes a responder to score the same question stated later on in the questionnaire in the same session, differently.

Summary of Findings

Hypothesis 1:

There is no unanimous agreement among all respondents on the sentiment score of a single response to any question within the survey.

This hypothesis is proven to be true. Not a single question had unanimous agreement by all the respondents. That is to say, not a single question had 100% positive, 100% neutral, or 100% negative.

Further, Appendix 1 demonstrates, at a glance, that the range of responses for all questions was 2. There was not a single answer that generated unanimity among the respondents. For every question, there was somebody who thought the statement was positive and somebody else who thought the statement was negative.

Hypothesis 2:

No two completed surveys will contain identical responses. Every completed survey will be unique.

Through the use of python programming, each of the 102 questionnaires received were compared against one another to see if there were any had identical responses. In other words, we used a program to identify if any two respondents in the sample mirrored respondent one's sentiment to the 25 questions. The result showed that no two sets of completed questionnaires were exactly alike amongst the sample investigated.

While sometimes respondents agreed on the overall Generic Sentiment Score (the sum of all the scores combined), how they arrived at that same score was different. In effect, the summation process effectively hides the underlining error. We do not view this result as the cancelling out of error. The error very much exists, and while the average Generic Sentiment Score was 4.21 the standard deviation was 5.39. In other words, 68.2% of the answers fell within a range of -1.18 and 9.60.

Hypothesis 3:

The reported mood of the scorer and the Generic Sentiment Score are correlated such that negative moods generate negative scores.

We did not find that as strong of a relationship as we had expected.

We ran the regression, and the R-squared value, which measures the predictive reliability between mood and score, was 0.055. This is a very low value, and while there is a general slope, the error is too high to draw a definite relationship. Bad moods don't seem to generate bad Generic Sentiment Scores (Appendix 2 and 3).

Hypothesis 4:

If the respondent starts off rating the first five statements as negative, the subsequent responses will tend to be more negative than those who started off positive.

There were no respondents who gave a negative sentiment score for all of the first five statements. As such this hypothesis could not be directly tested.

We analyzed the respondents who answered negatively to the first question only. Through the sum of scores, the respondents who answered negatively to the first question, generally also rated subsequent sentiments negatively.

"I'm working on learning excel/macros are there any really good books out that you'd recommend?"

Six out of 102 respondents rated this 'negative', an assigned score of -1.

Out of the five respondents we found that all of the surveys had a negative Generic Sentiment Score, ranging from -2.00 to -6.00 while the remaining responses in the sample tended to be much more positive (Appendix 4 and 5).

For these five people, starting out negatively is correlated with on the whole scoring it negatively. We can't argue a definitive cause and effect dynamic here, so we set this claim aside as not proven.

Hypothesis 5:

The tendency to anchor and adjust causes a responder to score the same text stated later on in the questionnaire, differently.

As some respondents noticed, Q2 and Q18 were indeed the same (DONE!! Tax Spreadsheet is done...now have to send it off to the accountant. Trying not to look at how much I spend on books). While 81% of respondents were consistent and scored that piece of text the same way twice, 19% of the sample did not (Appendix 6).

Of the 19% who provided different ratings to the same question, a further investigation was conducted to see if the question prior (the anchor) had any influence to the way a respondent rated the text.

We compared Q1/Q2 and Q17/Q18 together to investigate the tendency to anchor and adjust. In doing so we found a very weak relationship existed between these two sets of questions. Therefore, while the prior question plays somewhat of a role in how the following sentiment is rated, it is not major (Appendix 7 and 8).

It is telling though that on a panel of 102 humans, when examining the same statement twice, the result was only 81% accurate. The claim of 80% to 85% accuracy in machine sentiment analysis does indeed seem to be popular. Is it possible that a large group of humans themselves can only be 85% accurate.

Implications

The study proves the inaccuracies that exist with sentiment scores among human scorers. Hypothesis 1 proved that there was not unanimous agreement on the sentiment score of a single response to a survey question. Furthermore, Hypothesis 2 emphasized that overall responses to each questionnaire yielded different responses. There was no unanimity.

While the study investigated other influences on sentiment scores, such as mood and anchor and adjust, we found that these factors played a very weak role in impacting the sentiment score. However, the anchor and adjust had a stronger influence on scoring when compared to mood. Furthermore, the literature review mentions that sentiment scores often overlook language differences.

If a group of 102 humans could not agree, how can a single machine output a single score that everybody would agree with?

In sum, the Generic Sentiment Score appears inaccurate to most of us in part because humans themselves score differently. This is a very different form of instrumentation error than we are accustomed to in marketing, and poses significant challenges.

Recommended Alternatives to the Generic Sentiment Score

While it is easy to point out flaws in the existing paradigm, what do we propose instead?

We have to be cognizant that humans and especially machines will have accuracy limits. Even if there is a breakthrough and somebody wins the Turing Prize, you as the reader of this paper, will be a source of error. Your perception will more often than not diverge from popular opinion or the average truth score.

We suggest, instead, a more pragmatic approach.

We know that Generic Sentiment Score is not really a proxy for the health of a brand. We posit, respectfully, that perhaps there might be upwards of 60 potential variables, over and above Generic Sentiment Score that could give us better predictiveness and actionability into the health of a brand. Different variables will be relevant to different brands, and we are pessimistic that there exists a single silver bullet. Even Satisfaction Scores probably won't save us now.

Universal access of the social media data means that if a marketers wanted competitive intelligence, they could use the yardstick they're applying to themselves against the competition. This effectively solves the common problem of benchmarking, though, there will be dissatisfaction because everybody will use a different definition based on what is relevant within a given industry.

We also need to be cognizant of what frequently happens to us as marketers. It is precisely because nobody can really interpret 60 variables that we migrate down to a single, frequently inadequate, metric. And yet, especially with something as complex as what the public thinks of you – and the specific effects what a marketer does on that public – selecting a single yardstick is probably too simplistic to be useful.

We suggest that the Generic Sentiment Score should gradually give way to Composite Sentiment Score, or, more understandably, the concept of Brand Health. Brand Health, in turn, could become a long run predictive variable of profitability – just as satisfaction surveys in the past have been correlated.

We take the pragmatic position that we, as an emerging industry, will have as many different versions of Brand Health as there were responses to 25 questions. Error will persist. We ought to quantify and understand the degree of error instead of denying it.

We suggest that each social media measurement analyst examine, carefully, what is salient to his or her brand and take a broad view. We suggest they look for cause and reinforcing effects within their model. We recommend constructing their own version of Composite Sentiment Analysis based on what is relevant and actionable in their own competitive context. And, if they do not have enough time, talk to people who do understand.

Secondary Analysis

The dataset, as promised, is made available and accompanies this paper. This affords many of you to engage in secondary analysis.

Another hypothesis worthy of investigating is the impact of spelling and grammatical errors on the overall negative rating. It would be interesting to replicate this study for different languages and see where, if any differences exist.

We are also curious if these 102 records could be run through a program to see if a panel of 3 or 5 humans would be able to replicate the Mean Generic Sentiment Score produced by all 102 records. This would assume that the Mean Generic Sentiment Score actually represents 'the truth' as seen by those 102 people: a topic we look forward to talking about through Twitter, Facebook, and our blogs.

It is very likely that all 102 participants had differences among themselves: gender, income, age, mother language, education, familiarity with social media, and so on. It is also likely that people had differences within themselves as they completed the survey: increased joy, confusion, annoyance, and so on. The broader question of whether the Generic Sentiment Score produced by this study is reliable and valid was not tested. We invite anybody to design an experiment using the same questions to test validity and reliability.

We also invite users of sentiment analysis services to run their algorithms against these 25 questions, and see how they rate against this sample.

Conclusion

The Generic Sentiment Score is a poor social media measure and should not be used as the only data source when making actionable recommendations. As observed in the study, no two questionnaire sets had identical responses, which makes it difficult to sum up one piece of text with an overall sentiment score. People vary. As a result, their subjective responses will vary. This is a source of error. When using the Generic Sentiment Score or developing a more robust Composite Sentiment Score, one should be mindful of that error and acknowledge it.

Sources

1. Amos Tversky, Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases ." Science (1974): 1124-1131.

2. Carrabis, Joseph. Reading Virtual Minds. Scotsburn, NS: Northern Lights Publishing, 2009.

3. Francois, Laurent. "Sentiment analysis crap in social media." 03 September 2009. Socialmediatoday. 31 January 2010 <http://www.socialmediatoday.com/SMC/121398>.

4. Gattiker, Urse E. "Sentiment analysis for online content: Honest?" 03 September 2009. ComMetrics: tools for benchmarking social media. 31 January 2010 <http://commetrics.com/articles/fails-validity-test/>.

Appendix

Appendix 1: Response range for Hypothesis 1

	N	Range
	Statistic	Statistic
Q1	102	2.00
Q2	102	2.00
Q3	102	2.00
Q4	102	2.00
Q5	102	2.00
Q6	102	2.00
Q7	102	2.00
Q8	102	2.00
Q9	102	2.00
Q10	102	2.00
Q11	102	2.00
Q12	102	2.00
Q13	102	2.00
Q14	102	2.00
Q15	102	2.00
Q16	102	2.00
Q17	102	2.00
Q18	102	2.00
Q19	102	2.00
Q20	102	2.00
Q21	102	2.00
Q22	102	2.00
Q23	102	2.00
Q24	102	2.00
Q25	102	2.00
Valid N (listwise)	102	

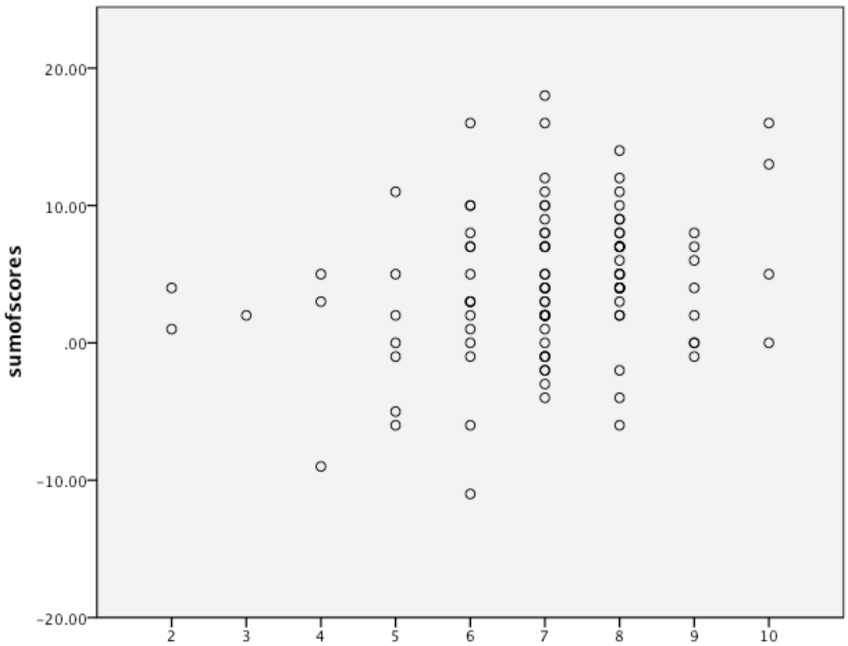
Range = positive, neutral and negative

Appendix 2: Generic Sentiment Score (Sum of Scores) for Hypothesis 3

sumofscores

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-11.00	1	1.0	1.0	1.0
	-9.00	1	1.0	1.0	2.0
	-6.00	3	2.9	2.9	4.9
	-5.00	1	1.0	1.0	5.9
	-4.00	2	2.0	2.0	7.8
	-3.00	1	1.0	1.0	8.8
	-2.00	3	2.9	2.9	11.8
	-1.00	6	5.9	5.9	17.6
	.00	6	5.9	5.9	23.5
	1.00	3	2.9	2.9	26.5
	2.00	11	10.8	10.8	37.3
	3.00	7	6.9	6.9	44.1
	4.00	10	9.8	9.8	53.9
	5.00	9	8.8	8.8	62.7
	6.00	2	2.0	2.0	64.7
	7.00	11	10.8	10.8	75.5
	8.00	6	5.9	5.9	81.4
	9.00	3	2.9	2.9	84.3
	10.00	5	4.9	4.9	89.2
	11.00	3	2.9	2.9	92.2
	12.00	2	2.0	2.0	94.1
	13.00	1	1.0	1.0	95.1
	14.00	1	1.0	1.0	96.1
	16.00	3	2.9	2.9	99.0
	18.00	1	1.0	1.0	100.0
	Total	102	100.0	100.0	

Appendix 3: R-Squared and Standard Error of the Estimate for Hypothesis 3



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.235[a]	.055	.046	5.26161

a. Predictors: (Constant), How would you describe your mood on a scale of 1 to 10 ?

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	162.227	1	162.227	5.860	.017[a]
	Residual	2768.450	100	27.684		
	Total	2930.676	101			

a. Predictors: (Constant), How would you describe your mood on a scale of 1 to 10 ?

b. Dependent Variable: sumofscores

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.661	2.479		-.670	.504
	How would you describe your mood on a scale of 1 to 10 ?	.835	.345	.235	2.421	.017

a. Dependent Variable: sumofscores

Appendix 4: The sum of the first five entries for Hypothesis 4

sumoffirstfive

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-3.00	2	2.0	2.0	2.0
	-2.00	8	7.8	7.8	9.8
	-1.00	18	17.6	17.6	27.5
	.00	36	35.3	35.3	62.7
	1.00	23	22.5	22.5	85.3
	2.00	12	11.8	11.8	97.1
	3.00	3	2.9	2.9	100.0
	Total	102	100.0	100.0	

Appendix 5: Frequency table of negative first responses for Hypothesis 4

negativenelly

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	97	95.1	95.1	95.1
	1.00	5	4.9	4.9	100.0
	Total	102	100.0	100.0	

Appendix 5: (cont'd)

sumofscores * negativenelly Crosstabulation

Count

		negativenelly		Total
		.00	1.00	
sumofscores	-11.00	1	0	1
	-9.00	1	0	1
	-6.00	1	2	3
	-5.00	0	1	1
	-4.00	1	1	2
	-3.00	1	0	1
	-2.00	2	1	3
	-1.00	6	0	6
	.00	6	0	6
	1.00	3	0	3
	2.00	11	0	11
	3.00	7	0	7
	4.00	10	0	10
	5.00	9	0	9
	6.00	2	0	2
	7.00	11	0	11
	8.00	6	0	6
	9.00	3	0	3
	10.00	5	0	5
	11.00	3	0	3
	12.00	2	0	2
	13.00	1	0	1
	14.00	1	0	1
	16.00	3	0	3
	18.00	1	0	1
Total		97	5	102

Appendix 6: Cross tabulation of Q2/Q18 for Hypothesis 5

V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books * V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books Crosstabulation

			V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books			Total
			-1.00	.00	1.00	
V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	-1.0	Count	30	5	2	37
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	75.0%	15.2%	6.9%	36.3%
	.0	Count	6	27	1	34
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	15.0%	81.8%	3.4%	33.3%
	1.0	Count	4	1	26	31
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	10.0%	3.0%	89.7%	30.4%
Total		Count	40	33	29	102
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	100.0%	100.0%	100.0%	100.0%

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.699	.072	9.502	.000
	Kendall's tau-c	.695	.073	9.502	.000
N of Valid Cases		102			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Appendix 6 (cont'd)

Statistics

V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

N	Valid	102
	Missing	0
Mean		-.059
Median		.000
Mode		-1.0

Statistics

V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

N	Valid	102
	Missing	0
Mean		-.1078
Median		.0000
Mode		-1.00

Appendix 7: Cross tabulation of Q1/Q2 for Hypothesis 5

I'm working on learning excel/macros are there any really good books out that you'd recommend? * V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books Crosstabulation

			V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books			Total
			-1.0	.0	1.0	
I'm working on learning excel/macros are there any really good books out that you'd recommend?	-1.00	Count	2	1	2	5
		% within V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	5.4%	2.9%	6.5%	4.9%
	.00	Count	10	16	16	42
		% within V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	27.0%	47.1%	51.6%	41.2%
	1.00	Count	25	17	13	55
		% within V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	67.6%	50.0%	41.9%	53.9%
Total		Count	37	34	31	102
		% within V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	100.0%	100.0%	100.0%	100.0%

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	-.187	.091	-2.060	.039
	Kendall's tau-c	-.168	.081	-2.060	.039
N of Valid Cases		102			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Appendix 7: (cont'd)

Statistics

V1. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

N	Valid	102
	Missing	0
Mean		-.059
Median		.000
Mode		-1.0

Statistics

I'm working on learning excel/macros are there any really good books out that you'd recommend?

N	Valid	102
	Missing	0
Mean		.4902
Median		1.0000
Mode		1.00

Appendix 8: Cross tabulation between Q17/Q18 for Hypothesis 5

this mofo greg just put his books in the freezer by accident and his parents thought it was funny. u might wanna check him for shrooms... * V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books Crosstabulation

			V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books			Total
			-1.00	.00	1.00	
this mofo greg just put his books in the freezer by accident and his parents thought it was funny. u might wanna check him for shrooms...	-1.00	Count	19	5	9	33
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	47.5%	15.2%	31.0%	32.4%
	.00	Count	17	23	11	51
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	42.5%	69.7%	37.9%	50.0%
	1.00	Count	4	5	9	18
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	10.0%	15.2%	31.0%	17.6%
Total		Count	40	33	29	102
		% within V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books	100.0%	100.0%	100.0%	100.0%

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	.217	.095	2.276	.023
	Kendall's tau-c	.208	.091	2.276	.023
N of Valid Cases		102			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Appendix 8: (cont'd)

Statistics

V2. DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

N	Valid	102
	Missing	0
Mean		-.1078
Median		.0000
Mode		-1.00

Statistics

this mofo greg just put his books in the freezer by accident and his parents thought it was funny. u might wanna check him for shrooms...

N	Valid	102
	Missing	0
Mean		-.1471
Median		.0000
Mode		.00

Appendix 9: Questionnaire

Experiment #1

Thanks for clicking on the link! The following experiment should take 5 to 8 minutes.

We want to share both the results and the dataset from this study openly.

At the end of this experiment, the results of this survey – the full CSV datafile - and accompanying paper will be published on syncapse.com, christopherberry.ca, and kevrichard.com. The authors are not keeping a record of anything personally identifiable. They are not asking for your name or email address.

Please indicate whether you believe the statements below are positive, neutral, or negative - towards books.

I'm working on learning excel/macros are there any really good books out that you'd recommend?

- Positive
- Neutral
- Negative

DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

- Positive
- Neutral
- Negative

Demonstrate Healthy Eating For Children : Play Library - Toys, Games, Books and Fun for Kids

- Positive
- Neutral
- Negative

Then i gotta go get books. ugh! its not that I'm lazy, its just that i dont like to do stuff!

- Positive
- Neutral
- Negative

Turfites_Tipple on "Another wage off our books": Remco van der Schaaf is set to sign a long-term loan deal with Da...

- Positive
- Neutral
- Negative

Pretty saddnd that no one RT my Iggy-Fund tweet to help me buy books. Guess noone cares bout my edcuation/future

- Positive
- Neutral
- Negative

It is a good thing for an uneducated man to read books of quotations. - Winston Churchill#quote

- Positive
- Neutral
- Negative

Going to buy some books! Yay for spending loads of $$$.

- Positive
- Neutral
- Negative

AHA naw i didnt. I watched shows and read books hahahaha. I went out though! To watch movies and stuff heh heh

- Positive
- Neutral
- Negative

Oh, my package is back in Fantoft...but do I still want it now :/ shame of the gifts in it and my books :(

- Positive
- Neutral
- Negative

@RandomName Highlight for Isaac: a Greco-Egyptian sarcophagus, as it had lions & elephants carved on it. Now, back to Star Wars & comic books.

- Positive
- Neutral
- Negative

RT @RandomName: We're still giving away books in our first 1000 followers draw. Prizes along the way. RT to win!

- Positive
- Neutral
- Negative

Taking a study break for a mani-pedi. Then it's back to the books for the big test tommorrow.
Trying not to freak out!

- Positive
- Neutral
- Negative

Just Chillin Got To Do Some Homework & Write That Letter So I Can Get My Books For The Semester It Was Rainin Today...

- Positive
- Neutral
- Negative

Off to the library to find some more Maeve Binchy books... Lowie, it's all your fault! ;)

- Positive
- Neutral
- Negative

@RandomName The world is full of art and books no one ever looked at. Makes good tinder, though. #litchat

- Positive
- Neutral
- Negative

this mofo greg just put his books in the freezer by accident and his parents thought it was funny. u might wanna check him for shrooms...

- Positive
- Neutral
- Negative

DONE!! Tax Spreadsheet is done... now have to send it off to the accountant. Trying not to look at how much I spend on books

- Positive
- Neutral
- Negative

via Twitter, the news, magazines, school, artists, books, etc. are all HUMAN FIRST! We make mistakes and we experience fear and struggle

- Positive
- Neutral
- Negative

I have now got 4 books on the iphone apps and I am delighted to say they are going well. If you have a book you would like on let me know

- Positive
- Neutral
- Negative

Anne Rice's Vamp. Chronicle books are great. Interview W/ A Vampire is the first book. the movie had brad pitt & tom cruise (meh)

- Positive
- Neutral
- Negative

In these books::craving something sweet::suggestions?

- Positive
- Neutral
- Negative

1 hour: free webinar on how to use JibberJobber (CRM) to market books. With one of my authors-all invited

- Positive
- Neutral
- Negative

Took girls to the school, read 80 pages of The Lovely Bones--doubt I can finish before Rita books arrive though.

- Positive
- Neutral
- Negative

RT @RandomName: I finallly finished reading The Host,I think it might have been the best book I've ever read.SM needs to write more books

- Positive
- Neutral
- Negative

How would you describe your mood on a scale of 1 to 10 ?