Formalizing your Curiosity

Wisconsin Public Library Consortium (WPLC) Final Workshop

ANALYSIS

Joshua H. Morrill, PhD
Morrill Solutions Research
Joshua H. Morrill, PhD
Morrill Solutions Research

Josh has been providing research and evaluation consulting with corporate, non-profit, and academic groups for the last 13 years. He worked as an emerging trends analyst for Gartner where he helped uncover some early usage of cell phones for game play.

As a consultant, Josh has worked extensively with the National Science Foundation in the assessment of various STEM (Science Technology Engineering & Math) technology initiatives. He has worked with National Science Digital Library project to assess the effectiveness of this initiative for faculty and librarians across the nation. And, more recently, Josh has worked with NSF and partners at the University of Illinois and University of Central Florida to conduct a large national study to develop a learner segmentation based on what people want out of learning technologies.

Josh has also partnered extensively with libraries. In particular, Josh has partnered with the Wisconsin Public Library Consortium (WPLC) in the development and implementation of a longitudinal study of Wisconsin public library users (and non-users) in 2003, 2007 and 2012. Josh also partnered with the University of Wisconsin and University of Illinois systems to assess student use and reactions to eTextbooks.

In addition to research, Joshua has also conducted several strategic planning and future-visioning workshops for corporate, library, education and non-profit groups that use scenario-based planning to help these organizations plan around future uncertainties. Joshua is a frequent speaker nationally. Josh holds a Ph.D. in communication science from the University of Wisconsin-Madison with a concentration in small group decision-making and quantitative research design and analysis.
Research is formalized curiosity. It is poking and prying with purpose.

-Zora Neale Hurston
ALL Measurement Has Error!

Some measures have less/ some more but it is always there

There is nothing known as "Perfect". Its only those imperfections which we choose not to see!!

—— Albert Einstein ——

The goal of research then is to minimize and parse (explain) the components that contribute to this error in order to make the best informed decision possible. Analysis is one way to do that.
What exactly do we mean by “Analysis”?

- Statistical Methods
  - Descriptive Methods
    - Graphs
    - Numeric Summaries
  - Inferential Methods
    - Confidence Intervals
    - Significance Testing
  - Coorelations
You have data...Now what?
Steps for approaching your data

1. Ask yourself
   How generalizable is this information? Is the sample sufficiently diverse/large?

2. Do some basic frequencies of your data?
   For every question frequency, mean, standard deviation

3. Ask yourself if any significance testing is needed
   Confidence Intervals? Mean tests?

4. How do I represent what I have found?
   What is the story you want others to take away.
The phenomena we are trying to measure is the earthquake.

The amount of people/cases we have determines is our seismograph is more or less sensitive.

Too FEW cases and we can only pick up on the BIG earthquakes.

Too MANY cases and every minor bump registers as an earthquake.
Profile of Respondents

**AGE**

- Less than 18 years old: 0.3%
- 18-24: 6.3%
- 25-31: 8.7%
- 32-38: 6.3%
- 39-44: 8.2%
- 45-51: 13.5%
- 52-58: 13.6%
- 59-64: 11.6%
- 65-70: 11.1%
- More than 70 years old: 20.2%

**MARITAL STATUS**

- Single, Never Married: 15.6%
- Married: 59.7%
- Divorced: 10%
- Separated: 0.4%
- Widowed: 10%
- Living with a partner: 4.4%

**PRESENCE OF CHILDREN**

- Younger than 5: 7.8%
- 5-12 years old: 13.2%
- 13-18 years old: 10.4%
- 18+ still living at home: 8%
- No children at home: 61.7%

**EMPLOYMENT**

- Full-time (not self-emp): 41.1%
- Part-time (not self-emp): 7.7%
- Unemployed: 2.7%
- Self-Emp/Bus Owner: 7.1%
- Homemaker: 4%
- Student: 2.9%
- Retired: 31.2%
- Other: 3.3%

**POLITICAL AFFILIATION**

- 30% Democrat
- 20% Republican
- 29% Not affiliated
- 15% Declined to answer
- 6% Other

**VOTING BEHAVIOR**

- 81% consider themselves a regular voter
- 17% consider themselves a non-voter
- 2% unsure
You have data...Now what?
Steps for approaching your data

1. Ask yourself
   How generalizable is this information? Is the sample sufficiently diverse/large?

2. Do some basic frequencies of your data?
   For every question frequency, mean, standard deviation

3. Ask yourself if any significance testing is needed
   Confidence Intervals? Mean tests?

4. How do I represent what I have found?
   What is the story you want others to take away.
Measures of Central Tendency

• Mean ... the average score

• Median ... the value that lies in the middle after ranking all the scores

• Mode ... the most frequently occurring score
Measures of Central Tendency

Mean ... the most frequently used but is sensitive to extreme scores

\[ e.g. \; 1 \; 2 \; 3 \; 4 \; 5 \; 6 \; 7 \; 8 \; 9 \; 10 \]

Mean = 5.5 (median = 5.5)

\[ e.g. \; 1 \; 2 \; 3 \; 4 \; 5 \; 6 \; 7 \; 8 \; 9 \; 20 \]

Mean = 6.5 (median = 5.5)

\[ e.g. \; 1 \; 2 \; 3 \; 4 \; 5 \; 6 \; 7 \; 8 \; 9 \; 100 \]

Mean = 14.5 (median = 5.5)
Variation or Spread of Distributions

Measures that indicate the spread of scores:

• Range

• Standard Deviation
Variation or Spread of Distributions

Standard Deviation

• It tells us how much the scores in the data set vary around the mean

• Low standard deviations indicate data clustered close to the mean, high standard deviations indicate the data are spread over a range of values.

MEANS ARE MEANINGLESS WITHOUT A STANDARD DEVIATION.
Standard Deviation

Scores:
I. e.g. 9, 10, 11, 7, 13
   Mean = 10  (SD = 2.2)

II. e.g. 1, 10, 13, 13, 13
      Mean = 10  (SD = 5.2)
...and would support their public library in a vote if the opportunity arises.

Q: If these was a referendum, ballot initiative or bond measure favoring your local public library on the ballot, how do you think you would vote?

- 35.3% Definitely vote YES
- 32.4% Probably vote YES
- 26.3% May vote either way
- 3.6% Probably vote NO
- 2.4% Definitely vote NO

Over 2/3 of respondents would likely vote YES
Less than 10% of respondents would likely vote NO
What would students pay?

In pilot one price was the most significant concern among students, and it remains a chief concern. However, in looking at comparative pricing most students would pay only about half of what they would pay for a used paper textbook for the Courseload eTextbook. The one exception was the Engineering course. The Engineering students would be willing to pay (on average) the same for an eTextbook and a used textbook.

What would you pay for…?

<table>
<thead>
<tr>
<th>Pilot courses</th>
<th>A NEW hard copy paper textbook you used in this class.</th>
<th>A USED hard copy paper textbook you used in this class.</th>
<th>A copy of the eText you used in this class.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>$118.52</td>
<td>$71.22</td>
<td>$43.37</td>
</tr>
<tr>
<td>Educational Psychology</td>
<td>$98.38</td>
<td>$63.00</td>
<td>$29.00</td>
</tr>
<tr>
<td>Philosophy</td>
<td>$69.77</td>
<td>$44.84</td>
<td>$22.16</td>
</tr>
<tr>
<td>Engineering</td>
<td>$93.47</td>
<td>$35.33*</td>
<td>$35.32*</td>
</tr>
</tbody>
</table>

Note: In all mean dollar amounts listed above, Standard deviations were nearly identical. The standard deviation was approximately $10.
Correlation

Correlation is a strength of association (-1 to +1), but it is not predictive. We do not know the cause of the association. Still it can be very helpful.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voting Behavior</td>
<td>r = .49</td>
</tr>
<tr>
<td>Age</td>
<td>r = .57</td>
</tr>
<tr>
<td>Gender</td>
<td>r = .63</td>
</tr>
</tbody>
</table>

Public Library Supporter
“You can say there is a correlation between rain and umbrellas opening.....But you would never say umbrellas caused the rain.”

-C. Halaby (UW-Madison)
Correlation

Correlations are useful but they are very susceptible to outlier data points.

Scatterplot 1

\[ r = 0 \]

Scatterplot 2 (with red outlier)

\[ r = .80 \]
You have data...Now what?
Steps for approaching your data

1. Ask yourself
   How generalizable is this information? Is the sample sufficiently diverse/large?

2. Do some basic frequencies of your data?
   For every question frequency, mean, standard deviation

3. Ask yourself if any significance testing is needed
   Confidence Intervals? Mean tests?

4. How do I represent what I have found?
   What is the story you want others to take away.
Confidence Intervals Handout

Formula for C.I. around a mean:

\[ \bar{x} \pm z \frac{s}{\sqrt{n}} \]

- \( \bar{x} \) = mean estimate
- \( N \) = sample size
- \( S \) = Standard Deviation of the mean
- \( Z \) = Desired confidence interval (95% 90%...)

This gives an upward bound and a lower bound to the mean estimate. So, if we have a sample with a mean of 8.5, a standard deviation of 1.25, a sample of 1,000 and we want to be 95% confident in the interval we derive, we get the following:

\[ 1.96 \times \left( \frac{1.25}{\sqrt{1000}} \right) = .08 \]

We add and subtract this to the mean: \( 8.5 + .08 = 8.58 \)  
\( & \) \( 8.5 - .08 = 8.42 \)

So we can say that we are 90% confident that the “true” estimate falls between 8.42 and 8.58.
Confidence Intervals to Compare Means

Using the previous example interval of 8.42 to 8.58, let’s look at two other datasets that measured the same construct (i.e. satisfaction)

<table>
<thead>
<tr>
<th>Comparison Set 1:</th>
<th>Comparison Set 2:</th>
</tr>
</thead>
<tbody>
<tr>
<td>X= 6.5</td>
<td>X= 8.25</td>
</tr>
<tr>
<td>N= 300</td>
<td>N= 100</td>
</tr>
<tr>
<td>S= 2.5</td>
<td>S= 1.5</td>
</tr>
<tr>
<td>Z= 90%</td>
<td>Z= 90%</td>
</tr>
<tr>
<td>Upper CI = 6.78</td>
<td>Upper CI = 8.54</td>
</tr>
<tr>
<td>Lower CI = 6.22</td>
<td>Lower CI = 7.96</td>
</tr>
</tbody>
</table>

In the first example at the right, we see that 6.78 -6.22 is an interval that does not overlap with 8.42 – 8.58. So we can say with 95% certainty that these intervals are significantly different.

In the second example, we see that 7.96 – 8.54 DOES overlap with 8.42 – 8.58. So we can say with 95% confidence that there is no difference between these two findings.
Confidence Intervals Around Percentages

\[ P\% \pm Z \left[ P\% \left(1 - P\%\right) / \sqrt{N} \right] \]

So, if we want to put a 95\% CI around a finding that 40\% of 275 people rated the taste of a new cola as “Very Good”:

\[
1.96 \times \left[ .40 \left(1-.40\right) / \sqrt{275} \right] = .40(.60)/\sqrt{275} = .24/16.58 = .015 \]

\[
1.96 \times 0.15 = .029
\]

.029 +/- .40

We add .029 to .40 and get upper value = .429 / We subtract from .40 and get a lower value = .371.

So we can now say that we are 95\% confident that the “true” percentage is between 37\% and 43\%.
You have data...Now what? Steps for approaching your data

1. Ask yourself
   How generalizable is this information? Is the sample sufficiently diverse/large?

2. Do some basic frequencies of your data?
   For every question frequency, mean, standard deviation

3. Ask yourself if any significance testing is needed
   Confidence Intervals? Mean tests?

4. How do I represent what I have found?
   What is the story you want others to take away.
Visualizations of data are quickly supplanting analysis in importance
Increasingly people want to be shown the conclusions—not told the conclusions.
Think about Contrast or Comparison Points

Educational Attainment – Comparison to U.S. Census Data

- Bachelors: Survey 49%, U.S. Census 69%
- Masters: Survey 28%, U.S. Census 26%
- Doctorate: Survey 23%, U.S. Census 5%
Think about Information Density (handouts vs slides)

1) Primary source materials to integrate into a course
   - Collection of Digital Resources
   - Search Engine (Google, Yahoo)
   - Faculty Overall: 0.9
   - Math Faculty: 0.6
   - Very Likely

2) Lesson plans, syllabi or exercises to integrate in instruction
   - Collection of Digital Resources
   - Search Engine (Google, Yahoo)
   - Faculty Overall: 0.43
   - Math Faculty: 1.23
   - Very Likely

3) Information about how to improve teaching
   - Collection of Digital Resources
   - Search Engine (Google, Yahoo)
   - Faculty Overall: 0.21
   - Math Faculty: 0.06
   - Very Likely

4) Education/pedagogy information
   - Collection of Digital Resources
   - Search Engine (Google, Yahoo)
   - Faculty Overall: 0.37
   - Math Faculty: 0.07
   - Very Likely

Very Unlikely
Across non-users, there appears to be an opportunity to increase awareness of what public libraries offer.
Visualizations work best to show complex (sometimes subtle) relationships

**LIBRARY PERCEPTIONS - % Top Box (Strongly Agree)**

<table>
<thead>
<tr>
<th></th>
<th>Essential services for children</th>
<th>Essential services for adults</th>
<th>In a bad economy the importance of libraries increases</th>
<th>Provide access to the information that I need</th>
</tr>
</thead>
<tbody>
<tr>
<td>I will am worse now than I was last year</td>
<td>70%</td>
<td>63%</td>
<td>62%</td>
<td>42%</td>
</tr>
<tr>
<td>I am better now than I was last year</td>
<td>62%</td>
<td>54%</td>
<td>58%</td>
<td>36%</td>
</tr>
<tr>
<td>Next year, I will be worse than I am now</td>
<td>71%</td>
<td>68%</td>
<td>69%</td>
<td>47%</td>
</tr>
<tr>
<td>Next year, I will be better than I am now</td>
<td>62%</td>
<td>56%</td>
<td>54%</td>
<td>33%</td>
</tr>
</tbody>
</table>
Visualizations work best to show complex (sometimes subtle) relationships

**LIBRARY PERCEPTIONS** - % Top Box (Strongly Agree)

- **Next year...I will be worse than where I am now**
- **I am worse now than I was last year**
- **I am better now than I was last year**
- **Next year...I will be better than I am now**

- **Public Libraries offer essential services for children**
- **Public libraries offer essential services for adults**
- **In a bad economy, the importance of a public library increases**
- **Public Libraries provide access to the information that I need**
Thank You

Joshua H. Morrill, PhD
joshua@morrillsolutions.com
608-588-2874