Quantitative Analysis

Robert B. Kvavik and Gail Salaway
ECAR Academy
Dedham, Massachusetts
May 17, 2006
How to Read Data

• ECAR strategy - KISS

• Methodology

• Data quality

• Measurement
Methodology

• Empirical analysis relies on or derives from observation or experiment
  – Empirical results can support a hypothesis
  – Guided by practical experience and not theory
  – ECAR studies use empirical analysis via surveys

• Quantitative analysis refers to the use of numerical and statistical techniques rather than the analysis of verbal material.
Methodology

• Qualitative analysis describes how individuals and groups view and understand the world
  – Meaning is constructed out of experiences.
  – Narrative-oriented
  – Uses content analysis methods
  – Some consider it’s goal is to generate hypotheses to be tested quantitatively.
Methodology

- **ECAR Quantitative Data:**
  - Survey of 18,039 freshman and seniors at 63 institutions, 2005

- **ECAR Qualitative Data:**
  - Analysis of comments of over 8,000 students to open-ended questions about their use of IT, 2005
Methodology

• Cross sectional analysis uses data that involves observation of a variable for some subsets of a population for the same point in time.

• Longitudinal analysis uses data that involves observations of a variable over a period of time.
# Cross Sectional Data

<table>
<thead>
<tr>
<th>Activity</th>
<th>N</th>
<th>Senior</th>
<th>Freshman</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating, reading, sending e-mail</td>
<td>17,865</td>
<td>99.7%</td>
<td>99.7%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Writing documents for your coursework</td>
<td>17,902</td>
<td>99.1%</td>
<td>98.7%</td>
<td>98.9%</td>
</tr>
<tr>
<td>Surfing the Internet for information to support your coursework</td>
<td>17,936</td>
<td>98.7%</td>
<td>98.1%</td>
<td>98.4%</td>
</tr>
<tr>
<td>Class activities and studying using an electronic device</td>
<td>17,961</td>
<td>96.4%</td>
<td>96.0%</td>
<td>96.2%</td>
</tr>
<tr>
<td>Surfing the Internet for pleasure</td>
<td>17,925</td>
<td>94.7%</td>
<td>95.0%</td>
<td>94.8%</td>
</tr>
<tr>
<td>Using a library resource to complete a course assignment</td>
<td>17,960</td>
<td>88.8%</td>
<td>86.9%</td>
<td>88.0%</td>
</tr>
<tr>
<td>Creating, reading, sending instant messages</td>
<td>17,782</td>
<td>74.2%</td>
<td>89.7%</td>
<td>81.1%</td>
</tr>
<tr>
<td>Downloading or listening to music or videos/DVDs</td>
<td>17,891</td>
<td>68.2%</td>
<td>83.8%</td>
<td>75.1%</td>
</tr>
<tr>
<td>Online shopping</td>
<td>17,905</td>
<td>77.2%</td>
<td>65.3%</td>
<td>71.9%</td>
</tr>
</tbody>
</table>

*Source: ECAR Study of Students and Information Technology, 2005*
## Longitudinal Data

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Firewall</td>
<td>71.4%</td>
<td>46.6%</td>
<td>24.8%</td>
<td>53.2%</td>
</tr>
<tr>
<td>Remote authentication dial-in user service (RADIUS)</td>
<td>54.4%</td>
<td>41.6%</td>
<td>12.8%</td>
<td>30.8%</td>
</tr>
<tr>
<td>Internet Protocol Virtual Private Network (IP VPN)</td>
<td>47.8%</td>
<td>33.0%</td>
<td>14.8%</td>
<td>44.8%</td>
</tr>
<tr>
<td>128-bit Wired Equivalency Privacy (WEP)</td>
<td>34.5%</td>
<td>33.4%</td>
<td>1.1%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Kerberos</td>
<td>21.2%</td>
<td>12.2%</td>
<td>9.0%</td>
<td>73.8%</td>
</tr>
<tr>
<td>Extensible Authentication Protocol (EAP)</td>
<td>19.7%</td>
<td>14.8%</td>
<td>4.9%</td>
<td>33.1%</td>
</tr>
<tr>
<td>40-bit Wired Equivalency Privacy (WEP)</td>
<td>19.6%</td>
<td>24.4%</td>
<td>-4.8%</td>
<td>-19.7%</td>
</tr>
<tr>
<td>Advanced encryption standard (AES)</td>
<td>14.2%</td>
<td>6.3%</td>
<td>7.9%</td>
<td>125.4%</td>
</tr>
</tbody>
</table>

*Source: IT Security in Higher Education, to be published in 2006, ECAR*
IT Security Taxonomy

- Electronic signature
- Perimeter firewalls
- Intrusion prevention
- Security standards for application development
- Shibboleth

- VPN for remote access
- Enterprise directory
- Interior firewalls
- Intrusion detection
- Active filtering
- Centralized data backup
- Intrusion prevention

Rate of Growth

Market Penetration

Low

High

Low

High

5/18/2006
Validity

• A *valid* measure is:
  – One which measures what it is supposed to measure.
  – Gets results that accurately reflect the concept being measured.
  – Meets reliability criteria
How Valid is the Data?

• Institutional information from official sources:
  • Carnegie Class 2000, Public vs. Private

• Institutional information from survey respondents:
  • Is your institution compliant with HIPAA? Yes/No/Don’t know

• Respondent information:
  • What is your preference for IT in a course?
How Valid is the Data?

Preference for IT in course

Source: ECAR Study of Students and Information Technology, 2005
Reliability

• Reliability: Accuracy of the scores of a measure.
  – Validity includes reliability; but,
  – Reliability does not imply validity

• Yardsticks might measure houses reliably, but not measure the lengths of insects reliably.
Confidence Level and Margin of Error

• How well a sample represents the whole population is gauged by:
  • Confidence Interval
  • Confidence level
  • Margin of error
Respondent Characteristics

- 34% are male; 66% female
- 44% are freshmen; 56% seniors
- 75% have a B or better GPA
- 13% are from BA institutions; 37% from MA institutions; 49% from doctoral institutions; and 1% AA and other

We achieved a 99% level of confidence with a +/- .02 margin of error.

What does this mean?
Confidence Level

• Confidence Interval (CI): *Range of values* within which the true population parameter lies
  - E.g. a Confidence that the percent of male senior and freshman responders in the true population (which is actually unknown) lies somewhere in the range of 33.98% to 34.02%

• Confidence Level: The *probability* associated with the CI
  - E.g. a 99% confidence level that the number of male responders in the true population is actually in that range of 33.98% to 34.02%

• Can be done only with competent sampling
Margin of Error

• Expresses the amount of random variation underlying a survey’s results (e.g. between +.02% and -.02%)

• The measure of variation in reported percentages if the same poll were to be repeated multiple times.

• States pollsters’ confidence in the data they have collected
Variables

• A *variable* is something that can assume a set of *values*.
  – GPA, age, attitudes, and opinions.
  – The values can be numerically represented
  – Variables are prone to variation
Units of Analysis

• Units of analysis can be many things
  – Individuals:
    • What is your title?
  – Departments (Central IT, Local Units):
    • Do campus departments at your institution control their own IT networks?
  – Colleges/universities:
    • At my institution, the IT governance process is effective.
      Scale = (1) strongly disagree to (5) strongly agree
They All Speak for the Institution

IT Governance Process is Effective
N=1850 IT professionals

Scale: 1 (Strongly Disagree) to 5 (Strongly Agree)

<table>
<thead>
<tr>
<th>IT Professionals</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior IT leaders</td>
<td>3.69</td>
</tr>
<tr>
<td>Report to Central IT</td>
<td>3.02</td>
</tr>
<tr>
<td>Report to Admin. Unit</td>
<td>2.88</td>
</tr>
<tr>
<td>Report to Academic Unit</td>
<td>2.68</td>
</tr>
</tbody>
</table>

Source: Information Technology Leadership in Higher Education, ECAR 2004
5/18/2006
The Power of Data: Scaling

• Nominal data: Categorical data where the order of the categories is not significant.
  – No/Yes/Don’t know answers to questions
  – Lists of academic disciplines, job titles, functional areas

• Ordinal data: Variables whose permissible values can be ordered and counted.
  – Characters are ordinal because we can call A the first character, B the second, etc.
  – Decimal numbers are not ordinal. If I say 1.0 is the first value, what is the second? If I say 1.1, then what is 1.01?
  – Example: Is your institution an (1) early adopter, (2) mainstream adopter, (3) late adopter?
Using Nominal and Ordinal Data

• Simple counts and frequencies
  – number of men (100) and women (50) in a sample (N=100)

• Percentages
  – percentage of men (67%) and women (33%) in a sample (total=100%)

• Proportions
  – 1/5 of the survey respondents reported having a disaster recovery plan for networking.

• Rank data on a scale of unequal intervals
  – Very poor, poor, fair, good, very good
## Reporting Nominal Data

<table>
<thead>
<tr>
<th>Position of person in charge of IT Security</th>
<th>Percent 2005</th>
<th>Percent 2003</th>
<th>Percent change</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT security officer (or equivalent)</td>
<td>34.9%</td>
<td>22.4%</td>
<td>12.5%</td>
</tr>
<tr>
<td>CIO (or equivalent)</td>
<td>14.3%</td>
<td>6.7%</td>
<td>7.6%</td>
</tr>
<tr>
<td>Director of administrative computing</td>
<td>2.7%</td>
<td>3.2%</td>
<td>-0.5%</td>
</tr>
<tr>
<td>Director of academic computing</td>
<td>1.2%</td>
<td>1.8%</td>
<td>-0.6%</td>
</tr>
<tr>
<td>Other academic management</td>
<td>0.6%</td>
<td>1.2%</td>
<td>-0.6%</td>
</tr>
<tr>
<td>Other administrative management</td>
<td>0.6%</td>
<td>3.2%</td>
<td>-2.6%</td>
</tr>
<tr>
<td>Other IT management</td>
<td>23.9%</td>
<td>30.9%</td>
<td>-7.0%</td>
</tr>
<tr>
<td>Director of networking</td>
<td>21.8%</td>
<td>30.6%</td>
<td>-8.8%</td>
</tr>
<tr>
<td>Total</td>
<td>100.0%</td>
<td>100.0%</td>
<td></td>
</tr>
</tbody>
</table>

Source: Information Technology in Higher Education, to be published in 2006, ECAR
# Reporting Ordinal Data

<table>
<thead>
<tr>
<th>Activities</th>
<th>N</th>
<th>Hours/week used on average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excluding cell phones, hours each week using an electronic device</td>
<td>17964</td>
<td>11-15 hours</td>
</tr>
<tr>
<td>Course activities and studying using electronic device</td>
<td>17281</td>
<td>3-5 hours</td>
</tr>
<tr>
<td>Writing documents for your coursework</td>
<td>17701</td>
<td>3-5 hours</td>
</tr>
<tr>
<td>Creating, reading, sending instant messages</td>
<td>14421</td>
<td>3-5 hours</td>
</tr>
<tr>
<td>Creating, reading, sending e-mail</td>
<td>17811</td>
<td>1-2 hours</td>
</tr>
<tr>
<td>Surfing the Internet for pleasure</td>
<td>16996</td>
<td>1-2 hours</td>
</tr>
<tr>
<td>Surfing the Internet for information to support coursework</td>
<td>17652</td>
<td>1-2 hours</td>
</tr>
<tr>
<td>Downloading or listening to music or videos/DVDs</td>
<td>13437</td>
<td>1-2 hours</td>
</tr>
<tr>
<td>Playing computer games</td>
<td>10836</td>
<td>1-2 hours</td>
</tr>
<tr>
<td>Using a course using course management systems</td>
<td>11356</td>
<td>1-2 hours</td>
</tr>
</tbody>
</table>

Source: ECAR Study of Students and Information Technology, 2005
The Power of Data: Scaling

- Equal interval data: Variables where distances between the values are equidistant
  - Years to degree, number of staff in the IT security unit.
  - Likert scale is ordinal, but often treated as Interval for analysis.
    "My institution is more secure than it was 2 years ago"
    Strongly disagree/disagree/neutral/agree/strongly agree

- Ratio data: Variables where there is a linear relationship between two quantities.
  - The ratio of 2:3 means 2 parts of one to 3 parts of the other.
  - A rate is a special kind of ratio where the two quantities being compared are of a different unit (miles per hour)
    "Rate of change in installing firewalls, 2003-2005"
Using Interval and Ratio Data

• Central tendency measures
  – Statistics that show how different units seem similar
  – Mean, median and mode

• Variability measures
  – Statistics that show how different units differ
  – Standard deviation of the sample
  – Range
Measures of Central tendency

• Mean: The average arithmetic value of set of numbers.

• Median: A number that separates the lowest-value half and the highest-value half of a sample or a population.

• Mode: The value that has the largest number of observations. In a bell curve distribution, the mode is at the peak.
Similar mean, median, mode

Source: ECAR Study of Students and Information Technology, 2005
CIO Bonuses in Higher Education, 2005

- $45,000 1 X
- $15,000 2 XX
- $10,000 2 XX
- $5,700 1 X
- $5,000 2 XX
- $3,700 4 XXXX
- $3,000 1 X
- $2,000 12 XXXXXXXXXXXXX

Mean = $6,088
Median = $3,000
Mode = $2,000
Measures of Variability

• Range: Smallest interval which contains all the data values.
  – Calculated by subtracting the smallest observation from the greatest
  – A poor measure of dispersion except when the sample size is very large (since it depends on only two of the observations).

• Standard Deviation: Measures the spread of the values.
  – The most commonly used measure of statistical dispersion
## IT Security Cultural Issues

<table>
<thead>
<tr>
<th>Issue</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual behaviors have become more sensitive to security and privacy in the past two years.</td>
<td>482</td>
<td>3.78</td>
<td>0.770</td>
</tr>
<tr>
<td>IT security architecture and implementation sacrifices some level of protection to ensure ease of use.</td>
<td>478</td>
<td>3.57</td>
<td>0.858</td>
</tr>
<tr>
<td>Business requirements take precedence over IT security when there is a conflict.</td>
<td>478</td>
<td>3.30</td>
<td>0.931</td>
</tr>
<tr>
<td>IT security inhibits academic freedom.</td>
<td>485</td>
<td>2.37</td>
<td>1.011</td>
</tr>
<tr>
<td>IT security compromises personal privacy.</td>
<td>480</td>
<td>2.16</td>
<td>0.902</td>
</tr>
<tr>
<td>IT security unnecessarily limits user access to information.</td>
<td>484</td>
<td>2.01</td>
<td>0.776</td>
</tr>
</tbody>
</table>

*Source: Information Technology Security in Higher Education, to be published in 2006, ECAR*
Behaviors more sensitive to security

Mean = 3.78  
Std. Dev. = .770

Security inhibits academic freedom

Mean = 2.37  
Std. Dev. = 1.011
Questions

kvavik@umn.edu

gsalaway@educause.edu