Harvard University
Program on Survey Research

Managing and Manipulating Survey Data:
A Beginners Guide

Chase Harrison
Department of Government
Harvard University

Types of Measures

- Interval / Continuous
  - Every possible value included

- Ordinal
  - All values can be placed above or below one another

- Nominal
  - Unique discrete categories

Types of Statistics

- Mean (average)
- Median
- Percentile
- Percentage

Types of Survey Questions

- Open-Ended
- Ordered Scales
- Discrete (yes/no)

Open Ended Questions

- "What do you think is the most important problem facing the country at the present time?"
- Data: "Well, it's mostly about unemployment. A lot of people don't have jobs. The war in Iraq is really bad too."

Coding Assigns Categories to the Responses

- Jobs/Unemployment/Lack of jobs/Looking for work
- The war/Iraq/Soldiers not coming home
- Terrorism/Homeland security/Protection from terrorists

Questions to Ask

- What about responses like:
  - "The economy"
  - "Terrorists blowing people up in Baghdad"
  - "Bush's silly war"
  - "The Democrats not supporting the war"

- Jobs/Unemployment/Lack of jobs/Looking for work
- The war/Iraq/Soldiers not coming home
- Terrorism/Homeland security/Protection from terrorists

  Depending on your hypotheses, you might combine new responses into existing ones or create new categories

The Next Step is Assigning Numbers to Data

- Code definitions:
  - Jobs/Unemployment/Lack of jobs/Looking for work
  - The war/Iraq/Soldiers not coming home
  - Terrorism/Homeland security/Protection from terrorists

- Code Labels can be shorthand descriptions of the fuller data
  - Unemployment
  - The Iraq War
  - Homeland Security

- Numbers often make nominal categories easier to manipulate
  - 1 Unemployment
  - 2 The Iraq war
  - 3 Homeland Security

Sometimes Open-Ended Questions can Be Pre-Coded

- What house (at Harvard) do you live in?
- What is your concentration?
- What state are you from?

  - *Often it's easier to ask these open-end than to provide all categories*

Closed-End Questions

Example:

"How likely do you think there will be a major terrorist attack in the next six months?"

❑ Extremely likely
❑ Very likely
❑ Somewhat likely
❑ Not very likely
❑ Not likely at all

Coding Turns this Into Numbers

1 Extremely likely
2 Very likely
3 Somewhat likely
4 Not very likely
5 Not likely at all
99 No response

Data

		Frequency
1	Extremely likely	59
2	Very likely	146
3	Somewhat likely	323
4	Not very likely	167
5	Not at all likely	39
99	Refused	18
	Total	752

Data:

		Frequency	Percent	Percent without nonresponse
1	Extremely likely	59	8%	8%
2	Very likely	146	19%	20%
3	Somewhat likely	323	43%	44%
4	Not very likely	167	22%	23%
5	Not at all likely	39	5%	5%
99	Refused	18	2%	
	Total	752		

Questions

- Should your percentages include or exclude people who say "don't know" from the base?
- Should your percentages include or exclude people who didn't answer the question from the base?

Discrete Categories

- Please indicate which of the following extracurricular activities you participate in…
  - ❑Political Organizations
  - ❑Cultural/Ethnic clubs
  - ❑Performing Arts Groups
  - ❑Varsity Athletics
  - ❑Intramural Athletics
  - ❑Other [Please Specify]

Coding Discrete Categories

- Each item can serve as its own measure…
  - Political Organizations (Yes/No/No Response)
  - Cultural/Ethnic clubs (Yes/No/No Response)
  - Performing Arts (Yes/No/No Response)

- One convention is to code these:
  - 1=Yes
  - 0 = No
  - 99 = Missing

Numbers

- Last week, how many times did you eat fish?
- 0 [Never/Didn't eat fish]
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8 or more
- Don't know

3

Numbers

- Can be asked directly or in groupings
- Think about special cases:
  - 0 / None
  - Doesn't apply
  - Maximum

Considerations

- In a data file, you may give a numeric code for "Don't know" or "Missing"
- Common codes:
  - 98 = Don't know
  - 99 = Refused / NA
- Negative codes are sometimes more useful
- Be sure not to include numbered codes for non-numeric responses in averages
- If a large portion of respondents answer the maximum category (e.g. 8 or more), means and other statistics might have problems

Using Categories for Numbers

- Approximately how many hours per week do you spend on academic work, outside of lectures or sections?

  - ❑ 0 - 10 hours
  - ❑ 11 – 20 hours
  - ❑ 21 – 30 hours
  - ❑ 31 – 40 hours
  - ❑ 41+ hours

Using Categories for Numbers

- Please estimate your total household income in 2006, before taxes…

  - ❑ Less than $25,000
  - ❑ Between $25,000 and $50,000
  - ❑ Between $50,000 and $75,000
  - ❑ Between $75,000 and $125,000
  - ❑ Between $125,000 and $200,000
  - ❑ Greater than $200,000

Coding Categorized Numbers

1 "0 - 10 hours"
2 "11 – 20 hours"
3 "21 – 30 hours"
4 "31 – 40 hours"
5 "41+ hours"
99 "No response"

1 "Less than $25,000"
2 "Between $25,000 and $50,000"
3 "Between $50,000 and $75,000"
4 "Between $75,000 and $125,000"
5 "Between $125,000 and $200,000"
6 "Greater than $200,000"
99 "No Response"

Attitude Scales

Please rate the following aspects of advising in your high school by placing one check for each aspect.

Advising Aspect	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Frequency of contact with advisor					
Advisor's knowledge of your schedule and requirements					
General availability of your advisor as a resource					
Adequate advisor to student ratio					

Attitude Scales

- 1 Very Dissatisfied
- 2 Somewhat Dissatisfied
- 3 Neutral
- 4 Somewhat Satisfied
- 5 Very Satisfied
- 97 Doesn't apply
- 98 Don't know
- 99 Refused/No Answer

Analyzing Attitude Questions

- Percentage
  - One category
  - Two Collapsed Categories

- Numeric
  - "Mean number"
  - Realize this is an ordinal mean
  - Numeric scale

Creating Scales from Multiple Questions

- Possible to create scales from multiple questions
- Can measure activities or attitudes
- Often treated as interval data
  - Mean or Median can be reported
- Sometimes scaled to 1, 10, or 100
- Reliability of scale should, ideally, be checkedcan be checked

Examples

- Please indicate which of the following extracurricular activities you participate in…
  - Political Organizations
  - Cultural/Ethnic clubs
  - Performing Arts Groups
  - Varsity Athletics
  - Intramural Athletics

Additive Scale

Respondent X….
  - 1 Political Organizations
  - 0 Cultural/Ethnic clubs
  - 1 Performing Arts Groups
  - 0 Varsity Athletics
  - 0 Intramural Athletics

Additive Scale: Number of activities
  - Scale for Respondent X: 2
  - This has meaning
  - "Seniors participated in an average of two different extracurricular activities, compared to freshmen, who participated in four activities."

Example of Attitude Scale

Attitude Scales

Please rate the following aspects of advising in your high school by placing one check for each aspect.

Advising Aspect	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Frequency of contact with advisor					
Advisor's knowledge of your schedule and requirements					
General availability of your advisor as a resource					
Adequate advisor to student ratio					

Data for Respondent X

- Contact   3   [Neutral]
- Knowledge   5   [Very Satisfied]
- Availability   1   [Very Dissatisfied]
- Ratio   3   [Neutral]

Interpreting Data

- Individual Items
  - Single Items
  - Collapsed

Interpreting Data

" Students were more likely to be satisfied with their advisors knowledge of requirements, compared to their advisor's availability. Eighty-five percent (85%) of students were satisfied with their advisor's level of knowledge, with 45% reporting they were extremely satisfied. This is in contrast with compared to only 65 percent of private school students. However, only 42 percent of students were satisfied with their advisor's availability, with only 20% being very satisfied."

Creating an Attitude Scale

- Make sure all answers are in same direction:
  - i.e. Positive attributes are scaled HIGH and negative attributes are scaled LOW
  - Often useful to make 0 the lowest category
  - Add questions together
  - Standardize scale so that 0 is minimum and 100 is maximum (scale is arbitrary convention)

Scale Data for Respondent X

- Contact   2   [Neutral]
- Knowledge   4   [Very Satisfied]
- Availability   0   [Very Dissatisfied]
- Ratio   2   [Neutral]

- Unstandardized Scale: 8
- Standardized Scale: 50
  - (Minimum = 0; Maximum = 16; Respondent=8)

Interpretation of Scale

- Best not to treat as if it is "real"
- Useful to compare means across respondents

Satisfaction with Advising by Type of School

	Satisfaction
Public High School	0.43
Charter High School	0.75
Private High School	0.80

Type of School