

Meeting 9

Measurement
(another example)

Sampling

Power

Measures

- Various fields have compiled sources of measures.
- E.g., <https://www.childwelfare.gov/topics/systemwide/assessment/family-assess/sources/>

Example: measure motives for volunteering

- The Functional Approach to Volunteers' Motivations, Gil Clary and Mark Snyder

The Motivations for Volunteering

<u>Values function</u>	the person is volunteering in order to express or act on important values, such as humanitarianism and helping the less fortunate
<u>Understanding function</u>	the volunteer is seeking to learn more about the world and/or exercise skills that are often unused
<u>Enhancement function</u>	the individual is seeking to grow and develop psychologically through involvement in volunteering
<u>Career function</u>	the volunteer has the goal of gaining career-related experience through volunteering
<u>Social function</u>	volunteering allows the person to strengthen one's social relationships
<u>Protective function</u>	the individual uses volunteering to reduce negative feelings, such as guilt, or to address personal problems

Volunteering, continued

Reasons for Volunteering

Using the 7-point scale below, please indicate how important or accurate each of the following possible reasons for volunteering is for you in doing volunteer work at this organization. Record your answer in the space next to each item.

	not at all important/ accurate for you	1	2	3	4	5	6	7	extremely important/ accurate for you
Rating									
—	1.								Volunteering can help me get my foot in the door at a place where I'd like to work
—	2.								My friends volunteer.
—	3.								I am concerned about those less fortunate than myself.
—	4.								People I'm close to want me to volunteer.
—	5.								Volunteering makes me feel important
—	6.								People I know share an interest in community service.
—	7.								No matter how bad I've been feeling, volunteering helps me to forget about it.
—	8.								I am genuinely concerned about the particular group I am serving.
—	9.								By volunteering, I feel less lonely.
—	10.								I can make new contacts that might help my business career.
—	11.								Doing volunteer work relieves me of some of the guilt over being more fortunate

Volunteering, continued

Volunteering Outcomes

Using the 7-point scale below, please indicate the amount of agreement or disagreement you personally feel with each statement. Please be as accurate and honest as possible, so we can better understand this organization.

- | | Strongly Disagree | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Strongly Agree |
|---------------|-------------------|--|---|---|---|---|---|---|----------------|
| <u>Rating</u> | | | | | | | | | |
| — | 31. | In volunteering with this organization, I made new contacts that might help my business or career. | | | | | | | |
| — | 32. | People I know best know that I am volunteering at this organization. | | | | | | | |
| — | 33. | People I am genuinely concerned about are being helped through my volunteer work at this organization. | | | | | | | |
| — | 34. | From volunteering at this organization, I feel better about myself. | | | | | | | |
| — | 35. | Volunteering at this organization allows me the opportunity to escape some of my own troubles. | | | | | | | |
| — | 36. | I have learned how to deal with a greater variety of people through volunteering at this organization. | | | | | | | |

Volunteering, continued

SCORING SHEET

VFI Career	Item	1	10	15	21	28	
	Response	___	+ ___	+ ___	+ ___	+ ___	= <u> </u> (SUM)

VFI Social	Item	2	4	6	17	23	
	Response	___	+ ___	+ ___	+ ___	+ ___	= <u> </u> (SUM)

VFI Values	Item	3	8	16	19	22	
	Response	___	+ ___	+ ___	+ ___	+ ___	= <u> </u> (SUM)

Outcomes Career (Add items 31 and 37) ___ + ___ =

Outcomes Social (Add items 32 and 38) ___ + ___ =

Outcomes Values (Add items 33 and 39) ___ + ___ =

Methods of collecting data

- Discussion example 1: NYC children age 6-12:
 - You want to survey/ treat / evaluate them.
 - But first you need to find them.
 - How?
- Discussion example 2: Maharashtra kids age 6-12:
 - Same or different approach to NYC?
 - Alternatives?

Sampling

- Major question: To whom do you want to generalize your results?
 - The larger group in which you are interested beyond the specific group at hand.
 - Critique of program evaluation is often at this point, i.e., external validity -- what is the larger situation that your evaluation addresses?
- Group to whom you wish to generalize
 - = Population
 - I.e., group from which you draw sample
- Types of Sampling
 - Non-Probability
 - Probability

Sampling terminology

Example: quality of life of homeless

- Parameters: Summary Descriptor of Population
 - Young adult, homeless men (b/t ages of 18 and 34)
- Census: Entire Population
 - Based on parameters (all homeless men?)
- Sampling Unit: Unit of Population
 - Individual (Shelters?)
 - Elements
- Accessible Population (vs. Theoretical)
 - Homeless men, ages 18 – 34
 - Living in shelters in NYC and LA
- Sampling Frame: List of Accessible Population
 - Actual List: Shelter List
 - Procedural List: Every person visible on street at 9 am
 - Random Digit Dialing

Sampling and statistics

- Sample
 - Sub-group of the population
- Statistic
 - Summary description of a variable in sample
 - Used as estimate of population parameter
 - Summary description of a variable in the population
- Sampling error
 - Estimate of the degree of error expected because sample is not population
 - Based on probability theory

Sampling and statistics 2

- Figure out degree to which your statistic reflects the actual population parameter.
- Estimate that with sampling distribution
 - What happens as you repeatedly repeat the sampling process... an infinite number of samples
 - Hopefully converges on central value
 - Likely for sample mean to be close to pop mean
 - Most of the samples are there
 - But...by sheer (bad) luck of draw, could draw extreme sample:
 - Statistics corrects for that: under the null hypothesis assume zero treatment effect. Ask how likely are we to see an extreme draw?
 - p value of 0.05 = 5/100 might have drawn an extreme sample

Probability sampling

- Sampling error
 - Statistical chance that because you cannot take an infinite number of samples from population, might end up with a sample that doesn't accurately reflect population
 - Random
 - Polls: Margin of error of +/- 3%
 - Only about sampling error (random error)
 - Does not include SYSTEMATIC error
 - » Leading questions
 - » Not including cell phones in sampling frame
 - Least of our measurement worries!

Types of probability sampling

- Simple random sample
 - Each element has equal chance of being selected
 - Selection is random (using random number table)
- Systematic sampling
 - Choose random start, select every nth unit
 - Interval: $\text{Dividing Population/Sample size}$
 - Problem if list has a meaningful order
- Stratified random sampling
 - Stratify sample by carefully selected variables
 - Randomly sample from within each cell

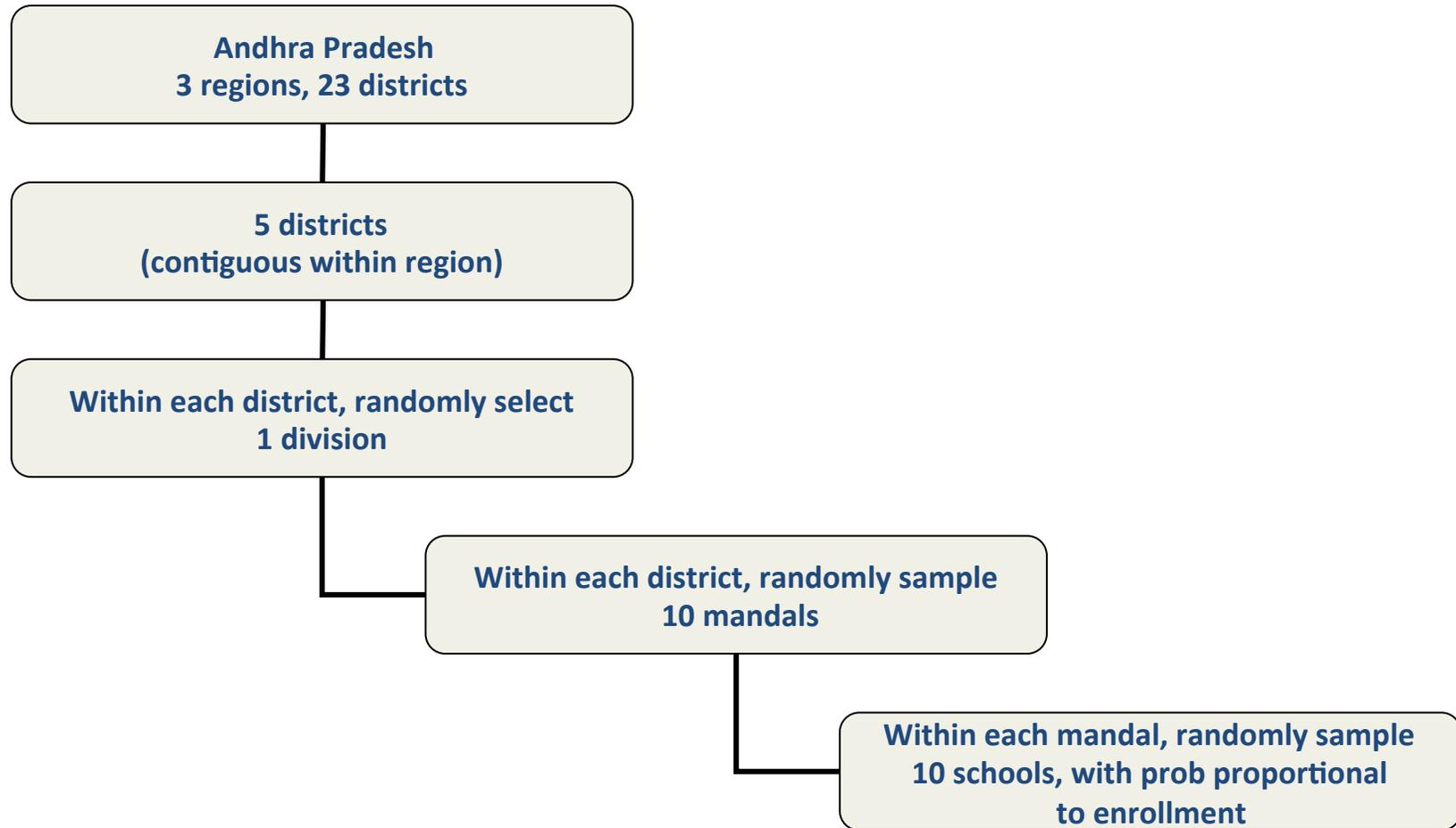
Types of probability sampling

- Cluster sampling
 - Randomly select clusters
 - Choose all elements within clusters
 - Greater sampling error
 - Units w/in cluster more homogenous than elements overall
- Multistage cluster sampling
 - Randomly select clusters
 - Randomly or systematically select elements within clusters
 - Students, charts, supervisors w/in clusters
- Disproportionate sampling (or over-sampling)
 - Give minority elements greater chance of selection
 - Over-sample for sub-group analysis
 - Weight up or down to reflect population

Non-probability sampling

- Convenience Samples - the name says it all.
- Quota Sampling
 - Convenience sample with an effort made to insure a certain distribution of demographic variables
 - Recruit participants within groups
 - E.g., Sample = 100. Quota =50/50 male and female
 - » Stop enrolling females once have 50
- Snowball Sampling
 - desired sample characteristic is rare, hard to locate respondents
 - relies on referrals from initial subjects to generate additional subjects
- Purposive
 - Selecting because of unique position of sampling elements
 - Executive Directors

Example: Muralidharan



Thus, the universe of 500 schools in the study was representative of the schooling conditions of the typical child attending a government-run primary school in rural AP.

Sample biases

- Non-probability sampling
 - Built in biases
 - Statistics less accurate
- Probability and non-probability sampling
 - Response rate!
 - Poor response rate turns probability sample into non-probability sample
 - Kerry Projections, 2004: Women
 - Large sample sizes don't solve the problem
 - If the 40% who drop out or refuse are extreme or similar
 - » Biased sample
 - Minimum? It depends...60%?

Example: attrition in KHDS

- A widely used survey in Tanzania used originally for health, but then later for child labor.

The data: baseline survey

- Baseline survey.
- Kagera Health and Development Survey (1991-1994)
 - 919 households, across Kagera region
 - 4 rounds between 1991/94

The data: 2004 resurvey

- 2700 households interviewed in 2004 (from baseline ~900).
- For 93% of the baseline households, at least 1 household member was re-interviewed in 2004.
- The KHDS 2004 tracked respondents outside of the village (within Kagera, Dar es Salaam, Mwanza, Uganda...). Tracking movers was critical. Without efforts to track children who moved out of the village, the recontact rates would have fallen from 81% to 49%.
- Attrition is inevitable, but a concern. Will presented results with imputation for missing values as a robustness check.

Re-interviewing the children ... 10-13 years later



The data: child labor and outcomes

- Child labor is total hours spent working in economic activities and household work in the previous 7 days. Economic activities are mainly farming and tending livestock.
 - Average of 16 hours per week (10 of which are in household work)
- Main outcomes:
 - Years of schooling: on average just over 6.
 - Completed primary: 78%.
- Other outcomes
 - Mobility: having moved away from village.
 - Farming.
 - Wage/salary work.
 - Marriage status.
 - Labor productivity.

Sample

- Sample is restricted to children who have not completed schooling in baseline and for whom child labor could affect schooling: those currently in school or those under 10 years and never enrolled.
 - 5,591 children in pooled cross-section: 1,971 children total.
 - 4,746 children in restricted pooled cross-section (in school or not yet enrolled): 1,747 children total, 75% of whom were re-interviewed in 2004.

Sample

- Of 1,747 children interviewed at baseline (1991-1994), 1,331 re-interviewed in 2004 (75%)
- Of 1,313 re-interviewed in 2004:
 - 694 still residing in original community
 - 637 moved to new village, new region, new country.
 - Half of these children moved more than 20 kilometers
- 36% nearby village
- 37% within region
- 22% new region
- 5% Uganda

Dealing with response bias

- Compare respondents w/ non-respondents
 - Usually on available socio-demographic data
 - On very relevant data
 - Longitudinal data where have previous data on outcomes before became lost to follow-up
 - Get contact data on non-respondents
 - Randomly sample and follow up aggressively
- Compare early and late respondents
 - Late respondents more like non-respondents

Power calculations to choose N

- Part of experimental / evaluation design: how many subjects?
- Recall that for a statistical test:
 - Type I error: Rejecting the null if the null is true
 - Type II error: Not rejecting the null if the null is false
- Size of a test is the probability of a type I error. Usually 0.05
- Power of a test is one minus the probability of type II error, i.e. the probability of rejecting the null if the null is false

Determine Power of Statistical Test

- Type I Error
 - Common choice 5% level of significance or Confidence = $1 - \alpha = 0.95$
 - I.e., you are willing to accept a 5% chance of false positive.
- Type II Error:
 - Common choice $\beta = 0.2$ or Power = $1 - \beta = 0.8$.
 - I.e., you are willing to accept 20% chance of saying there is no effect when there is, or a test that has 80% power (correctly categorizes a program as having no effect when that is true).
 - As you decrease Type I Error, increase Type II Error
- Seriousness of Type I vs Type II Error
 - Type I Error commonly seen as 4x as serious as Type II Error
 - 5/100 vs 20/100

HYPOTHESIS TESTING OUTCOMES		Reality	
		The Null Hypothesis Is True	The Alternative Hypothesis Is True
Research	The Null Hypothesis Is True	Accurate $1 - \alpha$ 	Type II Error β 
	The Alternative Hypothesis Is True	Type I Error α 	Accurate $1 - \beta$ 

Experimental Design: What affects power?

- The precision (i.e., $1/\text{variance}$) of within sample measures
 - Inherent variability of the outcome we are interested in.
 - Statistical power increases with the sample size in standard sampling.
- The size of the difference we are trying to detect.
 - Precision to \$1000 is a lot if we are trying to detect a 10k difference, little if we are hoping to detect a \$500 difference.
- But when is a sample “large enough”?
- We want to find N such that we will be able to detect an average treatment effect of size δ or larger with high probability.

Effect sizes

- Degree to which phenomenon in question actually exists (the program effect)
- Most common measure:
 - ES index $\frac{\text{Mean Group A} - \text{Mean Group B}}{\text{Std Dev}}$
 - *Difference between means expressed in std devs*

Effect sizes example

- Small
 - $D=.2$ (means differ by 2/10 of std dev)
 - Mean difference in height between 15 and 16 yr old girls (1/2 inch)
- Medium
 - $D=.5$ (means differ by ½ of a std dev)
 - Mean difference between 14 and 18 year old girls/women (1 inch)
- Large
 - $D=.8$ (means differ by almost a whole std dev)
 - Mean difference between 13 and 18 year old girls/women (3 inches)

Power in BB/BS study

Grossman-Tierney

- RA to Control ($n_1=472$) or Program ($n_2=481$)
- GPA Results
 - Control $X=2.63$ Program $X=2.77$
 - Pooled Std Dev=.40
 - $ES = .20 \rightarrow$ SMALL EFFECT
 - With actual sample size, power 0.9997
 - If $n = 200$, POWER = .94; if $n_1=n_2=129$, power of 0.8 is achieved.
- Skipped School
 - Control $X=4.5$ days Program $X=2.2$ days
 - Pooled Std Dev=3.8 days
 - $ES = .61 \rightarrow$ BIG EFFECT
 - If significance = 0.05, POWER = $>.99$
 - Sample size required to have .80 power = 43/group

Mortality before and after 2003 invasion of Iraq

Roberts, Lafta, et al.

- Case-Control Cohort Study
 - Every Cluster of Households Matched to Itself Before and After the Invasion
- Cluster Sample
 - Cluster=Households
 - Every household had an equal chance of being selected
- Sampling Procedures
 - Sample size determination (n=4,300)
 - Assumed a crude mortality rate of 10/1,000 per year (and SD of 107?)
 - Wanted 95% confidence (alpha=.05)
 - 80% power to detect a 65% increase in mortality requires sample of 4300.

Iraq mortality Study (cont'd)

- Every household assumed to have 7 individuals
 - Sample of 30 clusters of 30 households each
 - N=6,300
- Selected 33 clusters as target number (in case some too unsafe)
 - Iraq has 18 governates
 - Allocated number of clusters to sample to governates based on population size
 - Baghdad: target of 7 clusters vs. 1 cluster in other Governates
- Randomly assigned clusters to individual communities
 - List of communities in each governate, assigned random number to each
 - Selected number of communities based on target number of Clusters for that Governate
 - e.g., 100 communities, 3 communities randomly selected

Map



Figure 1: Crude mortality per 1000 people per year, by Governorate, before and after the invasion. Bar graphs represent number of deaths per 1000 person-years. Governorate rates are on a scale of 15 deaths per 1000 person-years, except for Anbar governorate, where deaths were more than ten times higher.

Iraq mortality study (cont'd)

- Now had to figure out how to identify clusters and then sample households within cluster
 - In randomly selected community
 - Drove to edges of area and stored coordinates in GPS
 - Assumed population lived within a rectangle formed by GPS
 - Drew area as a map subdivided by increments of 100m
 - » Randomly selected pair of #s between 0 and 100
 - » Went to that coordinate on map (15 and 34)
 - » Recruited nearest 30 households to be in study
 - ½ of 1% refused (5/988 households)
 - households clustered – corrected for homogeneity

Iraq morality study (cont'd)

- Actual Sample
 - 33 locations visited
 - 988 households chosen (7868 residents)
 - 99.9% response rate
- Results
 - Mortality rate prior to invasion: 5/1,000 per year
 - Mortality rate after invasion: 12.3/1,000 per year
 - Fallujah included in random sample, extreme outlier
 - If excluded, mortality rate after = 7.9/1,000
 - 98,000 deaths associated with invasion

Iraq – Time Series of Deaths

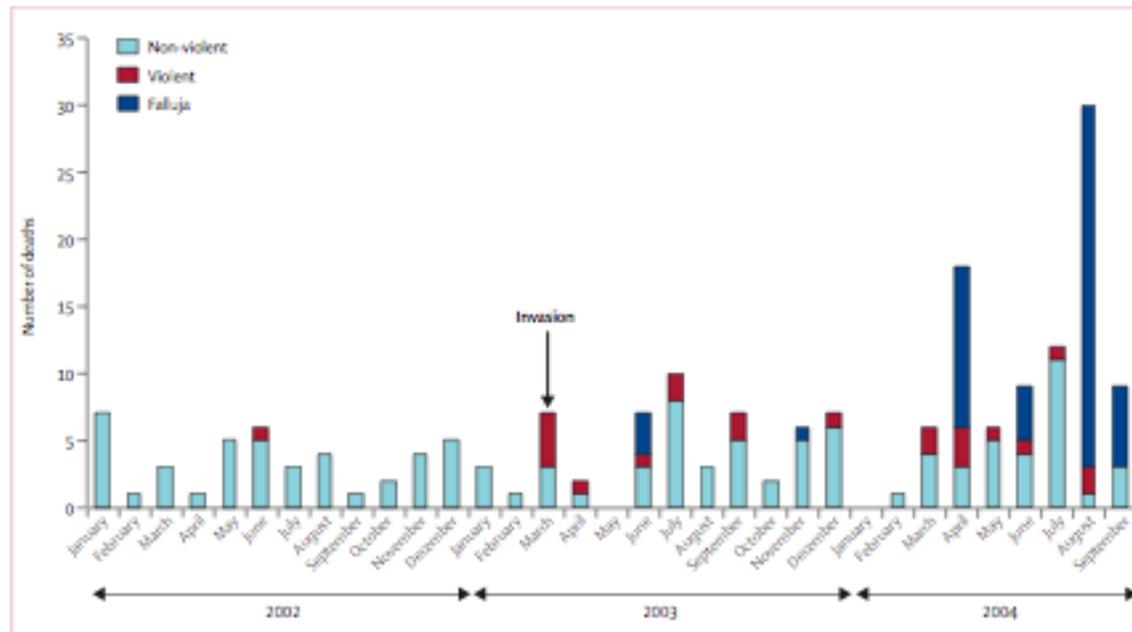


Figure 2: Number of deaths reported between January, 2002, and September, 2004