Module Objectives

By the end of this module participants will be able to:

1. Describe when to use quantitative methods
2. Identify 3 of the most commonly used quantitative data collections methods
3. Describe the strengths and weaknesses of commonly used quantitative data collection methods
4. Identify 3 frequently used sampling methods
5. Describe how and when to use descriptive statistics
6. Identify when some of the most common methods for inferential statistics should be used
7. Describe the differences between cost effectiveness analysis and cost benefit analysis and how to structure them
Using Quantitative Methods?

What are some of the advantages and disadvantages to using quantitative data?

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>❑ Standardized approach</td>
<td>❑ Limited information</td>
</tr>
<tr>
<td>❑ Easy to summarize</td>
<td>❑ Lack of context understanding</td>
</tr>
<tr>
<td>❑ Facilitates comparisons</td>
<td>❑ Up front planning time/expense if survey</td>
</tr>
<tr>
<td>❑ Possible to generalize</td>
<td>❑ Know precise information needed</td>
</tr>
<tr>
<td>❑ Possible to estimate magnitude and significance of differences</td>
<td>❑ Difficulty with sensitive questions and hidden populations</td>
</tr>
</tbody>
</table>
Frequently Used USAID Quantitative Tools and Methods (C-21)

1. In your experience, what are the most frequently used data collection methods for USAID evaluations?

2. Which are quantitative methods? What makes them quantitative?
Common Quantitative Data Collection Methods

Record Retrieval and Document Analysis

- Instrument developed to collect information to be extracted from records/documents
- Examples of sources
- Types of data extracted
- Advantages
- Major threats to validity

Example of a DCI

Extracted Data From Case Records for Sylhet District for 01 January – 31 December 2007

<table>
<thead>
<tr>
<th>Client Name</th>
<th># ANC Visits</th>
<th># TT Admin. During Pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ________</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Common Quantitative Data Collection Methods

Secondary Analysis of Existing Data Sets

- Often household survey data
- Advantages
- Major threats to validity

<table>
<thead>
<tr>
<th>Table 1: Key Indicators Bangladesh, Demographic and Health Survey 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicators</strong></td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>Neonatal mortality rate</td>
</tr>
<tr>
<td>Place of delivery</td>
</tr>
<tr>
<td>Skilled attendant at birth</td>
</tr>
<tr>
<td>At least 1 prenatal visit with skilled health provider</td>
</tr>
</tbody>
</table>

ACCESS: Safe Motherhood and Newborn Care. April 2008
Frequently Used for PEs and IEs: Sample Survey (C-20)

- Administered questionnaire
- Can be one-time or repeated; often time-series
- Use mostly close-ended items
- Yield quantitative data on value and extent of indicator
Potential Reliability Issues of Surveys

For high reliability, survey must provide stable values for a given case, over varying conditions, regardless if the measurement is done at different times, by different people, etc.

Some sources of inconsistency?
Potential Validity Issues of Surveys

Validity: The survey procedure provides accurate (true) values— the data represent what is supposed to be measured

- Credible that respondents would have knowledge/experience to answer?
- Measures cover all relevant concepts and excludes those not related?
- Measures actual behavior and not what respondent wants you to believe?
Improving Survey Reliability and Validity

• Keep it short! [Administered? 20-30 minutes]
• Use a cover letter to explain purpose, how data will be used, how respondent was selected, whether anonymity will be maintained, etc.
• Have experts review draft survey instrument
• Back-translate
• Develop a survey protocol and train administrators
• Obtain a good sampling frame
• Pretest, pretest and pretest!
• Data collection monitoring/auditing
• Others?
Construction Guidelines

Improving Survey Reliability and Validity

• Use only 3-4 item formats, more get confusing
• Do not have too many skip patterns
• Limit open-ended questions
• Use simple, short sentences
• Balance response choices around a neutral point
• Make sure coding categories are mutually exclusive and all inclusive
• Avoid double-barreled questions
• Put negative scale responses first
• Include “other”, “don’t know” or “not applicable” options
• Avoid leading questions and loaded wording
• Ask easy to answer questions first
Exercise: Land O’ Lakes

C-20

What does this survey instrument do well?

How might you improve it?
Evaluation Team Expertise for Surveys

- Subject matter knowledge
- Knowledge of culture(s) of intended survey recipients
- Sampling knowledge
- Experience developing and pretesting surveys
- Knowledge of recommended procedure for translation, if applicable
- Knowledge and experience in developing survey protocols, and training and supervising of field administrators (when applicable)
- Knowledge and experience in data entry and data cleaning
- Knowledge and skills in quantitative analysis
Addressing 2 Common Threats to Validity

How to better ask recall questions

• Limit period of time, but 12 months too long; some say 1 day
• Clarify concepts such as “chronically ill” by giving examples using daily living activities
• Ask multiple questions
• Stimulate recall with specific events in time period
• Ask for specific events at time of interest

How to better minimize the socially desirable response

• Give a range (e.g. earnings)
• Read an intro on importance of accuracy
• Provide an intro that say alternatives are OK (e.g. reasons for not voting)
• Multiple question as to whether yesterday the behavior of interest was average, less or more than average, then asking for the #
• Compare to friends— more, less, about the same
Electronic Data Collection

What is it?
- Using PDAs, Smartphones, Netbooks, Ipads, etc to collect data

Why would we want to do this?
- Quick access to data
- Reduce transcription error
- Programmable skip patterns

Why not?
- High upfront costs
- Specific skills required for programming and operation
## Survey Development

### Paper v. Electronic

<table>
<thead>
<tr>
<th>Paper Surveys</th>
<th>Electronic Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pen and paper Sometimes called “PAPI” (Pen and paper interviewing)</td>
<td>PDAs, Tablets, Netbooks, Smartphones Sometimes called “CAPI” (Computer Assisted Personal Interviewing)</td>
</tr>
</tbody>
</table>

**Why:** Readability, Quick start-up, Respondent perception, Interviewer familiarity

**Why not:** Some errors easier (skip questions), data entry errors, long-term cost of printing, logistical burden

**Why:** Quick access to data, Reduce transcription error, Programmable skip patterns, Internal (real-time) consistency checks, Active prompts and alerts, Easy to implement survey experiments, Respondent perceptions, Cheaper long-run

**Why not:** High upfront costs, Time to program survey, Specific skills required for programming and operation, Interviewer familiarity, Equipment/electricity/network
Introduction to Probability Sampling

Objectives
1. Know 3 kinds of probability sampling used frequently in USAID
   a. Simple random sample
   b. Stratified random sample
   c. Cluster sample
2. Understand trade-offs of uncertainty and precision
3. Understand how to estimate sample size
### Sampling Definitions Mix & Match Quiz

<table>
<thead>
<tr>
<th>KEY TERMS</th>
<th>DEFINITIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Census</td>
<td>A. The type of entity on which data is sought (people, sites, neighborhoods, etc.)</td>
</tr>
<tr>
<td>2. Sample</td>
<td>B. The full set of units about which you want to infer conclusions</td>
</tr>
<tr>
<td>3. Unit of Analysis</td>
<td>C. Information collected from all units in a group</td>
</tr>
<tr>
<td>4. Population (Universe)</td>
<td>D. Data collected from some of the units in a group</td>
</tr>
<tr>
<td>5. Sampling Frame</td>
<td>E. A list of units of analysis used for drawing a sample from a population</td>
</tr>
</tbody>
</table>
Sampling Exercise

• Read through Section 1 (Sampling Method 1) of Exercise 14-1

• Consider the questions on your own and be ready to discuss
Probability Samples: Random Samples

• What is it?
  – Every unit has the same likelihood of being selected
  – With sufficient numbers, sample representative of the population
  – Require a comprehensive ‘sampling frame’

• Can make estimates about the larger population based on the subset-- cost effective

• Can rule out sample selection bias
Generating Simple Random Numbers

- Establish a sample size and proceed to use a random number table to select units until reaching the sample size.

- iPhone apps and online random number generators: e.g. RANDOM.ORG; stattrek.com

- On Excel, use =rand()
Probability Samples: Random Samples

• Are there potential drawbacks to simple random sampling?
  – What if we have small, but important, sub-groups?
  – What if our budget limits transport costs?

 ➢ Allows us to get representative data, minimizing selection bias, but often, more sophisticated versions of random sampling are preferred
Systematic Random Sampling

Random Start, Fixed Interval Sample

- Use when there is a sequential population that is not already enumerated and would be difficult and time-consuming to enumerate
- Select units from a list, but on the basis of a fixed interval after a random start
- If a sample of 50 is required from a population of 455 health workers, this means a sample of 50/455, or one in nine units is the fixed interval.
- A random number between 1 and 9 is used to select the 1st health worker. Then every 9th health worker is selected thereafter
- If the initial random number is 6, the selected health workers will be the sixth, the fifteenth, the twenty-fourth, etc
Stratified Random Samples

• What is it?
  – Divide population in groups (strata) sharing certain characteristic(s) and randomly select sample from within these groups

• Use when specific groups, usually small proportions of the total population, must be included that might otherwise be missed by using a simple random sample

• Two types: Proportional and Disproportional stratified random samples.
Cluster Sampling

• What is it?
  – Most population units cluster—teachers in schools, health workers in rural clinics, families in villages
  – *Single-stage cluster sampling* = clusters are randomly selected and every population unit in the selected cluster is included in the sample

• Many USAID evaluations use cluster sampling because there are no defined population or strata lists
  – Even if there were lists, population is too dispersed = $$$
Cluster Sampling: Example

• Survey of agricultural extension workers:
  – Project has 50 extension units = 50 clusters
  – Each cluster has 5 extension workers

  ➢ 10 clusters are selected through simple random sampling
     and all the extension workers in each selected cluster are
     interviewed
     ➢ Total sample of 50 workers

• This reduces transport costs (simple random sampling may
  have pulled one worker from each unit)

• BUT…comes at a price…Intra-cluster Correlation (ICC)
Cluster Sampling: Intra-cluster Correlation

• What is it?
  – The correlation of outcomes within a cluster
    • Also called intra-class correlation

• Why does this matter?
  – Thought experiment: Want to measure public opinion before upcoming election. Is it better to:
    • A: Sample 50 people randomly and interview on political views
    • B: Sample 5 households and interview all 10 members of each HH
  – Political views within the HH are highly correlated so we learn less about the population from each additional individual in a HH

• Intra-cluster correlation can significantly reduce our ‘effective’ sample size
Two-Stage or Multiple Stage Cluster Sampling

• What is it?
  – Extension of cluster random sampling, taking a random sample from the selected clusters

• Examples:
  – In the survey of agricultural extension workers, 10 clusters were selected, then 3 out of 5 extension workers in each were selected for interviews through simple random sampling
  – In a survey of 500 students:
    • Simple random sampling: 500 students randomly selected
    • Cluster random sampling: 10 schools randomly selected, all students in each selected school surveyed
    • 2-stage cluster random sampling: 10 schools randomly selected, 5 classes in each school randomly selected, all students in selected classes surveyed
    • What would 3-stage cluster random sampling look like?
Sample Size Basics

• What is it?
  – The number of units from which we seek to collect data

• **Remember:** we use data from our sample to *infer* or estimate information about the whole population
  – Estimate population proportion (X% of population satisfied with program)
  – Estimate population mean (average income of population)

• We often then use the estimates to compare population means
  – Before vs. After
  – Treatment vs. Control
How Large a Sample Do I Need?

First decision: Choosing the Confidence Interval

• Statistics estimate the probability that the sample results are representative of the population as a whole--always have a margin of error with samples = confidence intervals

• Expressed as a range: plus or minus some number-- for example, if 48% percent of your random sample selects an answer, +/- 4% had you had asked the question of the entire relevant population between 44% (48-4) and 52% (48+4) would have picked that answer

• By increasing sample size, you increase precision and decrease margin of error

• The larger the confidence interval, the less precise the results will be-- > ± 5 is considered reasonable
Second decision: How confident do I need to be?

- Evaluators must choose how confident they need to be—expressed as a percentage.
- The confidence level represents how often the true percentage of the population who would pick an answer lies within the confidence interval.
- Generally use 95% confidence level.
- Meaning: 95 times out of 100, sample results will accurately reflect the population as a whole.
- The higher the confidence level, the larger sample needed.
Research aims for a 95% confidence level and a ±5% or smaller confidence interval.

Again using the example “If 48% percent of your random sample selects an answer” you can make statements like:

- You are 95% sure that the true percentage of the population that would have chosen this response is between 44 and 52 because 48% percent of your random sample -4 and +4 is 44 and 52.

All else equal, what would happen to the confidence level if we decreased our confidence interval?
### Sample Sizes for Small Population = 100

<table>
<thead>
<tr>
<th>Confidence Interval</th>
<th>Confidence Level 99%</th>
<th>Confidence Level 95%</th>
<th>Confidence Level 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 1%</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>± 2%</td>
<td>98</td>
<td>96</td>
<td>94</td>
</tr>
<tr>
<td>± 3%</td>
<td>95</td>
<td>92</td>
<td>88</td>
</tr>
<tr>
<td>± 5%</td>
<td>87</td>
<td>80</td>
<td>73</td>
</tr>
</tbody>
</table>
## Sample Sizes for Large Population 1,000,000

<table>
<thead>
<tr>
<th>Confidence Interval</th>
<th>99%</th>
<th>95%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 1%</td>
<td>16,576</td>
<td>9,604</td>
<td>6,765</td>
</tr>
<tr>
<td>± 2%</td>
<td>4,144</td>
<td>2,401</td>
<td>1,691</td>
</tr>
<tr>
<td>± 3%</td>
<td>1,848</td>
<td>1,067</td>
<td>752</td>
</tr>
<tr>
<td>± 5%</td>
<td>666</td>
<td>384</td>
<td>271</td>
</tr>
</tbody>
</table>

IPDET 2009
## Sample Sizes by Population (95% +/- 5%)

<table>
<thead>
<tr>
<th>Population Size</th>
<th>Sample Size</th>
<th>Population Size</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>550</td>
<td>226</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>600</td>
<td>234</td>
</tr>
<tr>
<td>40</td>
<td>36</td>
<td>700</td>
<td>248</td>
</tr>
<tr>
<td>50</td>
<td>44</td>
<td>800</td>
<td>260</td>
</tr>
<tr>
<td>75</td>
<td>63</td>
<td>900</td>
<td>269</td>
</tr>
<tr>
<td>100</td>
<td>80</td>
<td>1000</td>
<td>278</td>
</tr>
<tr>
<td>150</td>
<td>108</td>
<td>1200</td>
<td>291</td>
</tr>
<tr>
<td>200</td>
<td>132</td>
<td>1300</td>
<td>297</td>
</tr>
<tr>
<td>250</td>
<td>152</td>
<td>1500</td>
<td>306</td>
</tr>
<tr>
<td>300</td>
<td>169</td>
<td>3000</td>
<td>341</td>
</tr>
<tr>
<td>350</td>
<td>184</td>
<td>6000</td>
<td>361</td>
</tr>
<tr>
<td>400</td>
<td>196</td>
<td>9000</td>
<td>368</td>
</tr>
<tr>
<td>450</td>
<td>207</td>
<td>50000</td>
<td>381</td>
</tr>
<tr>
<td>500</td>
<td>217</td>
<td>100000 +</td>
<td>384</td>
</tr>
</tbody>
</table>
Sample Sizes by Population

Relationship between the Size of a Population and a Sample with a 95% Confidence Level and 5% Margin of Error
Determining Sample Size

Most stat pages have a sample size calculator

[Image: Diagram showing the process of determining sample size]

Confidence Level: 95% or 99%
Confidence Interval:
Population:
Sample size needed:

www.surveysystem.com/sscalc.htm

www.researchinfo.com/docs/calculators/samplesize.cfm
Rapid Surveys/Mini-Surveys for PEs

• Mini-surveys or rapid surveys using cluster sampling often used for broad patterns, trends, and tendencies— not precise measurements
  – Sample sizes usually 25–70 cases
  – Remember ICC with cluster sampling
  – Use structured closed-ended questionnaires for quick analysis
  – 5-8 weeks total

• For example, in evaluating an agricultural project for small farmers, it is often immaterial if the beneficiary approval rating is 60 or 63 percent. But credibility may be an issue!

See (on the CD) Krishna Kumar “Conducting Mini-Surveys in Developing Countries” for more information. USAID.
Other Types of Sampling

• Is probability sampling the only way?
  – Typical Case, Maximum Variation, Quota Sample, Extreme Case, Confirming/Disconfirming Cases
  – Snowball
  – Convenience

• These are more commonly associated with qualitative methods where the focus isn’t as strong on generalizability

• When might these be most useful?
Using Statistics

Objectives
1. Know difference between descriptive statistics and inferential statistics
2. Know how to calculate measures of central tendency
3. Know how to calculate measures of dispersion
4. Understand when to use what type of inferential analysis
Two Types of Statistics

Descriptive

- Summarize numerical data
- Describe how many and what percentage of a distribution share a particular characteristic
  - 33% of the respondents are male and 67% are female

Inferential

- Generalize from a small set of data on a random sample to a larger set (population)
- Risk of error because a sample may be different from the population as a whole
### Descriptive Stats: Frequency Distribution

#### Numbers and Percentages of Farmers Purchasing Fertilizer By District

<table>
<thead>
<tr>
<th>District</th>
<th>Total Farmers</th>
<th>Year 1 N (%)</th>
<th>Year 2 N (%)</th>
<th>Year 3 N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10,000</td>
<td>3,612 (36)</td>
<td>4,170 (42)</td>
<td>4,670 (47)</td>
</tr>
<tr>
<td>B</td>
<td>14,000</td>
<td>765 (05)</td>
<td>1,241 (09)</td>
<td>2,073 (15)</td>
</tr>
<tr>
<td>C</td>
<td>38,000</td>
<td>21,036 (55)</td>
<td>20,217 (53)</td>
<td>19,416 (51)</td>
</tr>
</tbody>
</table>
Descriptive Statistics

Measures of central tendency

• How similar are the data? *How similar are the ages of the people in this group?*

• The 3Ms
  – **mode**: most frequent response
  – **median**: midpoint or middle value in a distribution
  – **mean**: arithmetic average

Measures of dispersion

• How different are the data? *How much variation in the ages exists?*

• Range, frequency distribution, and standard deviation (SD)
Measures of Central Tendency

- **Mode** = the observation that appears most frequently in the distribution, or the one with the largest number of responses. Most distributions have only one modal category. When two categories are nearly or equally prominent, the distributions are called bimodal.

- **Median** = a measure that divides the distribution into two equal parts. It is computed by locating the middle observation. For an odd number of cases, the middle item is calculated by adding 1 to N (the number of cases) and dividing the sum by 2. The median is easier to calculate for an odd number of cases than it is for even cases. For an even number of cases, the middle item is calculated by adding the middle two items and dividing by 2.

- The most widely used measure of the central tendency is arithmetic average or the **Mean**. It is calculated by adding all the values of the distribution and dividing the sum by the total number of cases.
Measures of Dispersion

- To know how different the data are, we need measures of the variation in the values of a distribution— the range, variance, and standard deviation.
- What happens to the graph if we reduce dispersion?
Range: Easy to Compute but Least Useful

- Simplest of the 3 measures
- **Range** = the difference between the highest and the lowest values of the distribution
- Compute by subtracting the lowest value from the highest
- Because the range depends on the two extreme scores, it is an unstable measure
Variance and Standard Deviation

- **Variance** = measure of the spread of the scores around the mean.

- **Standard deviation** = square root of the variance.

- In a ‘normal distribution’:
  - 68% of cases are within 1SD
  - 95% within 2 SD
  - 99% within 3 SD
To Calculate Mean, Variance, Standard Deviation

Enter all the numbers separated by comma ",,\".  
E.g: 13,23,12,44,55

Results

Total Numbers: 
Mean (Average): 
Standard deviation: 
Variance(Standard deviation):

NOTE: EXCEL and SPSS also will perform calculation
Performance Evaluations

Should calculate (1) measures of central tendency and (2) measures of dispersion

- frequency distributions
- range, mean, median
- standard deviations

...and sometimes basic inferential statistics
Inferential Statistics

- Used to analyze data from randomly selected *samples*

- Often compare 2 groups to determine if the differences between the groups occurred by chance variation in samples taken from the two populations or from something else (like the intervention!)

- Risk of error because sample may be different from the population as a whole

- IEs typically use inferential statistics

- PEs may use them to examine the difference between baseline and post-intervention measures on outcome indicators (before-after)
T-Test

• What is it?
  – How confident are we based on our sample data that the population is significantly different from:
    • A defined target (*Did we meet our target?*)
    • The mean of another population, based on a second sample (*Was T group different from C group?*)

• What data do we need?
  – Data on our variable of interest from:
    • One sample (if we are comparing against a target)
    • Two Samples
      – Can be same units: Before-After
      – Or Different: Treatment-Control
T-Test

• What are the results?
  – P-value:
    • The probability that the difference in sample means occurred due to chance (sampling)
    • In other words, the chance that the populations means are actually the same
    • How does this relate to confidence level?
      – Confidence level = 1 - P

• Confidence level of the test represents our confidence that the result is correct
  – If the confidence level is above our threshold (usually 95%), we say the result is ‘statistically significant’
T-Test

• Let’s see what this looks like in practice:
  – T-test can be done in Excel using ‘=t.test’ function or by adding the ‘Data Analysis’ add-in

• Note: When comparing 2 or more sample means… ANOVA (ANalysis Of VAriance)
Regression

• What is it?
  – Looks at relationships between variables
    • What is the relationship?
    • What happens to our DV if we increase/decrease the IV?
      – What is the relationship between education and life expectancy?
      – …or between our treatment and outcome of interest
    – Can we be confident that it isn’t due to sampling?

• What data do we need?
  – Data on our DV (or outcome) and our IV(s) (or explanatory variables)
What’s the relationship?
What’s the relationship?

Life Expectancy

Education
Regression

• What are the results?
  – P-Value: same interpretation as t-test
  – Coefficient: relationship between DV and IV
    • Is it positive or negative?
    • Is it large or small (but remember the units)?

• Can be extended to look at multiple variables
  – What is the relationship between two variables holding other variables constant?
    • E.g. How are education levels related to life expectancy, for individuals with the same income?
  – In this way, regression analysis mimics matching
Which Quant Method?

• Remember the BRIGHT program…

• Take a look at Exercise 14-2 and answer the questions with your group.
Evaluation Sample Size Planning

• When we want to compare sample means (before vs after or treatment vs control), sample size calculation is more complicated
  – Before we were just looking at one sample

• So how do we know what sample size we need in order to be confident in our conclusions?
  – **Power Calculations**
Power

• What is it?
  – The likelihood we will conclude our program had an impact of at least a given magnitude, assuming it actually had that effect
  – What are the chances our t-test will yield a significant result if the treatment and control populations are actually different?
  – If the program actually had the impact, why would we ever conclude that it didn’t?
    • It is possible that we by chance selected an unrepresentative sample
      – E.g. the ‘best’ from the comparison and ‘worst’ from the treatment

• Typically look for 80% or 90% power
Power Calculations

- Calculate the probability of correctly concluding an impact under a set of assumptions:
  - **Effect Size**: Size of program impact that we want to measure
    - Larger anticipated impacts are easier to measure (↑ Power)
  - **Sample Size**: Number of sampled units in treatment and comparison (combined)
    - Increased sample size = ↑ Power
  - **Significance Level**: Likelihood of a ‘false positive’
    - Usually set at 5% or 10%, higher = ↑ Power
  - **Variance**: Variation in the outcome of interest within T & C groups
    - Less variance = ↑ Power
  - **Intra-Cluster Correlation**: how much do units in a cluster resemble one another
    - Low variance within clusters significantly reduces effective sample size
Effect Size

• Minimum program impact that we want to be able to measure
  – Larger anticipated impacts are easier to measure (↑ Power)
  – Increasing effect size, increase power

• We won’t be able to distinguish smaller changes from ‘no effect’

• But can you increase your effect size?
  – Increased resources
  – Effect size is usually fixed by resources and technical approach
Effect Size

• How do you select an effect size?
  – Be careful using ‘expected’ impact
    • If the effect is any smaller, you will not be able to measure it
  – What is the smallest effect that would justify the program being adopted or continued?
    • If the effect is smaller than that, it might as well be zero

• Not interested in proving that a very small effect is different from zero
Sample Size

• Number of sampled units in treatment and comparison (combined)
  – Be careful if much more than 50% sample comes from T or C

• Increasing sample size makes us more confident that each sample represents its population
  – Increased sample size leads to increased power

• Sample size is typically what we have most control over in determining power
  – Although increasing sample size almost always increases data collection costs
    • Sometimes increases can be marginal
Power Calculation: Putting it all together

• Although it may sound complicated, it is most important to remember how each variable effects power and which variables can be influenced.

• Power Calcs should be done before a project driving decisions on sample size and design
  – STATA
  – Optimal Design:
    – WT Grant Foundation-Consultation Services and Optimal Design
    – Online power calculators:
      • http://www.dssresearch.com/toolkit/spcalc/power_a2.asp

• More of an art than science…be conservative in assumptions
What you really need to know about Power

• Most of the factors are pretty much set so check that the assumptions for these are reasonable
  – Ask in SOW to see power calcs

• Focus your attention on effect size and sample size
  – Don’t use ‘expected’ effect…think minimum useful/important effect
  – If the sample size is too large for the budget, don’t change assumptions
    • Consider different design or data collection methods
  – In general, there is no minimum/required sample size…
    • Can vary from dozens to tens of thousands
    • Although possible, n<100 should raise red flag

• Take a look at Exercise 14-3
Cost-Benefit Analysis and Cost Effectiveness Evaluation

Objectives

1. Understand the difference between cost-benefit analysis and cost-effectiveness evaluation
2. Understand how cost-benefit is calculated
3. Understand how cost-effectiveness is calculated
Cost-Benefit Analysis

• A cost-benefit ratio is determined by dividing the projected benefits of a program by the projected costs
• Infrastructure projects often compare the ERR forecast to that attained
• ERR: quantifiable net economic benefits to society
• USAID examples

• For social projects, a wide range of variables, including non-quantitative ones such as quality of life, are generally considered
Cost Effectiveness Analysis

- Type of economic evaluation that examines both the costs and outcomes of alternative competing intervention strategies

Often used for health interventions: Eg. compare a smoking prevention program targeted at adolescents with a smoking cessation program targeted at committed adolescent smokers in cost per life-year gained
Cost Effectiveness Analysis

Figure 1:
J-PAL COST EFFECTIVENESS: additional years of student attendance per $100 spent
www.povertyactionlab.org

- INFORMATION
- HEALTH INTERVENTIONS
- INCENTIVES/REDUCED COSTS
- MULTIPLE OUTCOMES

<table>
<thead>
<tr>
<th>Program</th>
<th>Years of Additional Attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on Returns MADAGASC.</td>
<td>28.6 years §</td>
</tr>
<tr>
<td>Deworming at School KENYA</td>
<td>3-4 years §</td>
</tr>
<tr>
<td>Iron &amp; Deworming INDIA</td>
<td>2.8 years §</td>
</tr>
<tr>
<td>School Meals KENYA</td>
<td>1.4 years §</td>
</tr>
<tr>
<td>Girls' Merit Scholarships KENYA</td>
<td>1 year</td>
</tr>
<tr>
<td>Subsidized Uniforms KENYA</td>
<td>.09 years §</td>
</tr>
<tr>
<td>CCTs for Secondary Ed. MEXICO</td>
<td>.02 years §</td>
</tr>
<tr>
<td>CCTs for Primary Ed. MEXICO</td>
<td></td>
</tr>
</tbody>
</table>
In Closing:

1. Qualitative methods (next module) help explain findings from quantitative methods

2. Not everything that is statistically significant is important

3. Many more research and learning materials can be found on the Course CD’s
Review Questions

• What are the 3 most common quantitative data collection methods?

• What are three frequently used sampling methods?

• What is the difference between cost effectiveness analysis and cost benefit analysis?