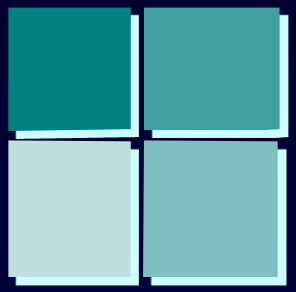
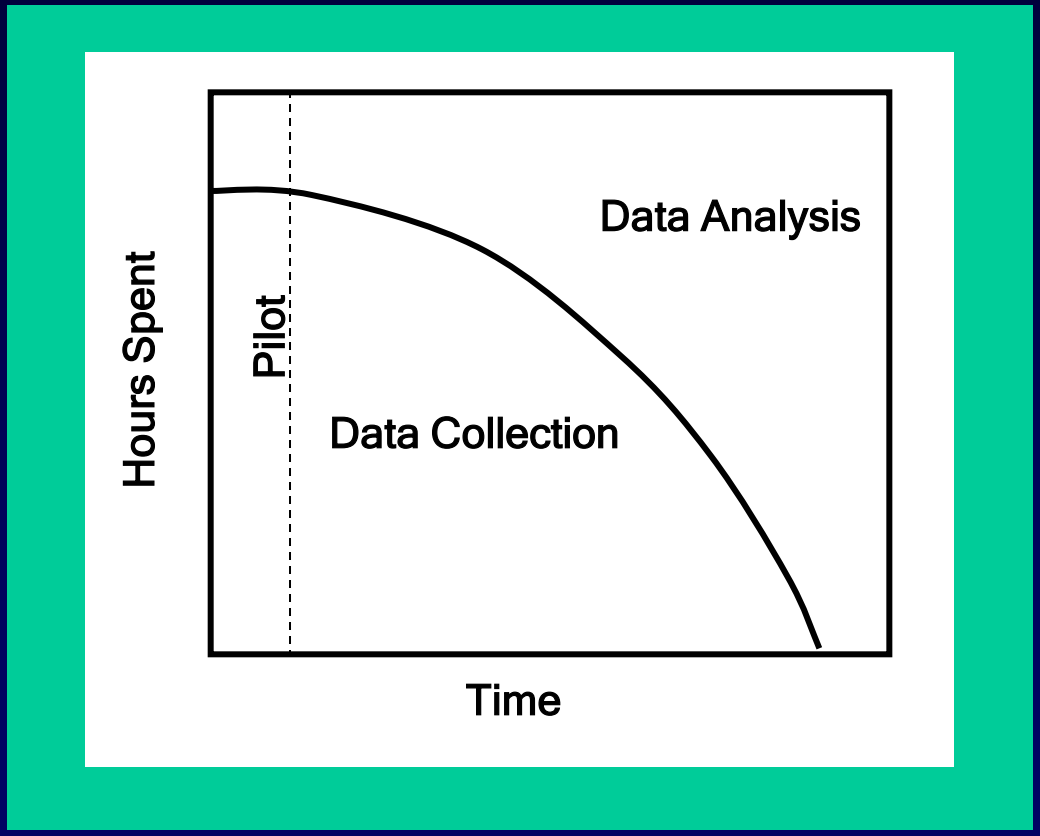


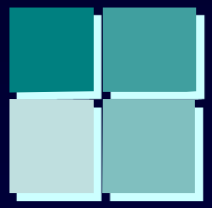
Module 10: Planning for and Conducting Data Analysis

Introduction

- Data Analysis Strategy
- Analyzing Qualitative Data
- Analyzing Quantitative Data
- Linking Quantitative Data and Qualitative Data

Data Collection and Analysis



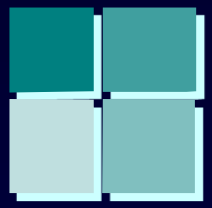


Qualitative Analysis

- Best used for in-depth understanding of the intervention
- Answers questions like:
  - What are some of the difficulties faced by staff?
  - Why do participants say they dropped out early?
  - What is the experience like for participants?

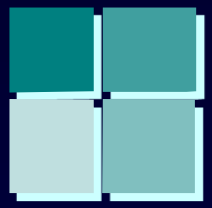
Contrast with Quantitative Analysis

- Used to answer questions like:
  - What are the mean scores for the different groups of participants?
  - How do participants rate the relevance of the intervention on a scale of one to five?
  - How much variability is there in the responses to the item?
  - Are the differences between the two groups statistically significant?



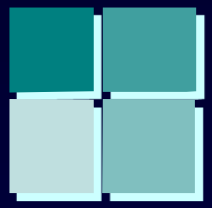
Tips for Collecting and Analyzing Qualitative Data

- Write up impressions, ideas, and interview or observation notes daily
- Progressively focus
- Meet frequently with team to compare notes and adjust
- Keep a file of quotations



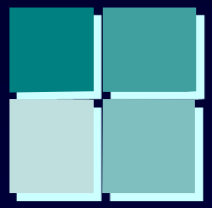
Analyzing Qualitative Data

- Nonnumerical data collected as part of the evaluation:
  - E.g. open-ended interviews, written documents, focus groups transcripts
- May use content analysis to identify themes and patterns
- Also progressive focusing-- ongoing analysis in which themes emerge



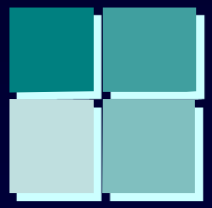
Drawing-out Themes and Patterns

- As you review, begin to make notes
- Goal is to summarize what you have seen or heard:
  - common words
  - phrases
  - themes
  - patterns



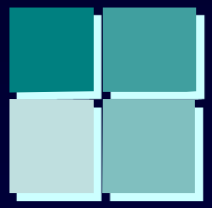
Iterative Process of Coding

Review, revise, redefine, add to and sometimes discard codes as field notes suggest more empirically driven labels



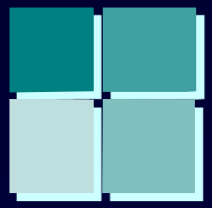
Content Analysis

- Identify certain words or concepts in text or speech
- Conceptual analysis:
  - look at word frequencies
- Relational analysis:
  - look at word frequencies
  - explore relationships among concepts



Computer Help for Qualitative Data Analysis

- Software packages to help you organize data
- Search, organize, categorize, and annotate textual and visual data
- Help you visualize the relationships among data
- Necessary when have large amount of data

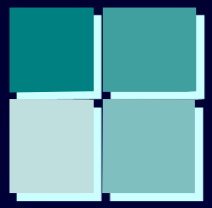


Examples of QDA Software

- OSR's N6 from QSR (formerly NUD*IST)
- Ethnograph
- Qualpro
- Hyperqual
- Anthropax
- Atlas-ti
- Nvivo 8
- AnSWR
- HyperRESEARCH
- Qualrus
- others

Manual Analysis of Qualitative Data

- Materials needed:
  - several highlighters (different colors)
  - a worksheet for each evaluation question
  - data, including notes, transcripts, and recordings from interviews or focus groups
  - collection tools for self-completed questionnaires, registration forms, observations, or chart reviews



Manually Coding Data

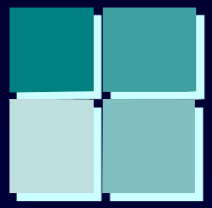
- Read all of the data carefully
- Come up with names or labels for topics, issues, or themes = codes
- Using codes, classify all of the data
  - cut with scissors to manually sort (copying first)
  - may use a number coding system
  - index cards may be useful

Example Qualitative Worksheet

Evaluation Question: Were participants satisfied with the training workshops?
Color, code, or symbol: Yellow

Topics	Quotes	Findings
Parents decide on topics 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ////	I think the process of deciding would be valuable.	There was a strong feeling that parents should be more involved in the choice of topics
Cover a couple of topics per session 𝍸 𝍸 𝍸 ///	Sometimes we just got into a topic and then it was time to leave or move to something else. We need more time to discuss.	Many participants (38 or 52 interviewed) thought there should be more time for discussion.
Not enough time spent on each topic 𝍸 𝍸 𝍸 ///		

Source: Porteous, et al 1997



Organizing and Interpreting Qualitative Data

- Develop categories
- Code the categories
- Check for reliability (more than one observer)
- Analyze the data
- Share and review information
- Write the report

Triangulation and Analysis

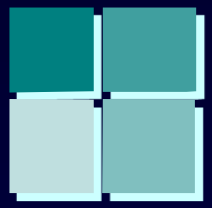
If evaluation has obtained data from three or more *sources of information*, analyze findings for consistency

E.g., program staff, government officials, beneficiaries

If evaluation has used three or more *data collection instruments*, analyze findings for consistency

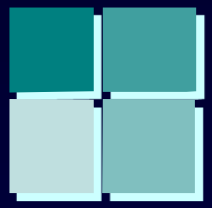
E.g., interviews, focus groups, questionnaires

questionnaires, existing data, expert panels



Concluding Thoughts on Qualitative Data

- Qualitative data collection is not the easy option
  - labor intensive and time consuming
  - reliability among coders, using a coding scheme is essential
- Can reveal valuable information

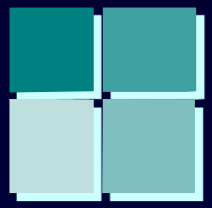


Quantitative Data: Using Statistics

- Quantitative data are numerical and analyzed with statistics
  - descriptive statistics: used to describe and analyze data collected about a quantitative variable
  - inferential statistics: used with random sample data by predicting a range of population values for a quantitative qualitative variable

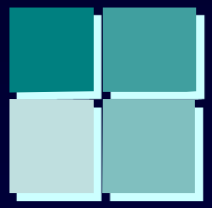
Coding Quantitative Data

- Coding allows data to be processed in a meaningful way
- Data need to be transformed into numeric responses. Examples:
  - yes coded as 1; no coded as 2
  - data collected in ranges; ranges given numbers
- Codes placed in data dictionary



Cleaning the Data

- Removing errors and inconsistencies
- Common sources:
  - missing data, blank responses, typing errors, incorrectly formatted data, column shift, fabricated data, coding errors, measurement and interview errors, out-of-date data



Descriptive Statistics

- Describes how many and what percentage of a distribution share a particular characteristic
- Example:
  - 33% of the respondents are male and 67% are female

Example of Descriptive Statistics in a Table

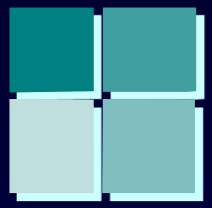
How many men and women are in the program?

Table 11.5: Distribution of Respondents by Gender

Male		Female		Total
Number	Percent	Number	Percent	Number
100	33%	200	67%	300

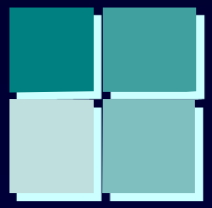
Source: Fabricated Data

Write up: Of the 300 people in this program, 67% are women and 33% are men.



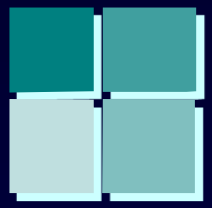
Distributions

- Measures of central tendency
  - how similar the data are
  - example: How similar are the ages of the people in this group?
- Measures of dispersion
  - how different the data are
  - example: How much variation in the ages?



Measures of Central Tendency

- The 3-M's
  - mode: most frequent response
  - median: midpoint or middle value in a distribution
  - mean: arithmetic average
- Which to use depends on the type of data you have
  - nominal, ordinal, interval/ratio

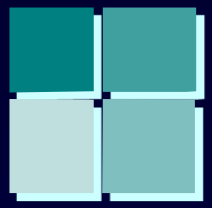


Nominal Data

- Data of names or categories
- Examples:
  - gender (male, female)
  - religion (Buddhist, Christian, Jewish, Muslim)
  - country of origin (Burma, China, Ethiopia, Peru)
- With nominal data, use *mode* as best measure of central tendency

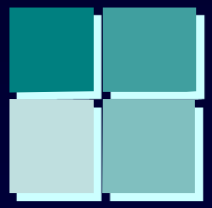
Ordinal Data

- Data that has an order to it but the "distance" between consecutive responses is not necessarily the same
- Lacks a zero point
- Examples:
  - opinion scales that go from "most important" to "least important" or "strongly agree" to "strongly disagree"
- With ordinal data, use ***mode*** or ***median*** as best measure of central tendency



Interval/Ratio Data

- Data of real numbers, numbers with a zero point and can be divided and compared into other ratio numbers
- Examples:
  - age, income, weight, height
- With interval/ration data, use *mode*, *median*, or *mean* as best measure of central tendency – the choice depends on the distribution
  - for normal data, *mean* is best
  - for data with few high - or - few low scores, *median* is best



Calculating

- Mode: the response given most often
- Median: place data in sequential order then count down to half way
- Mean: (most people think of it as the average)

Example Data

Table 11.7: Sample Data (fatalities as % of all landmine victims)

Country	% fatalities
Colombia	65
Angola	60
Croatia	41
Afghanistan	61
Cambodia	58
Sri Lanka	75

Source: Fabricated Data

Example Calculations for % Fatalities Data

- Mode: no mode, all have only one data point
- Median: total entries is 6, with data in order two middle scores are
  (61 and 60) ÷ 2 = 60.5
- Mean:

  (65+60+41+61+58+75) ÷6 = **60**

Measures of Dispersion

- Range
  - difference between the highest and lowest value
  - simple to calculate, but not very valuable
- Standard deviation
  - measure of the spread of the scores around the mean
  - superior measure, it allows every case to have an impact on its value

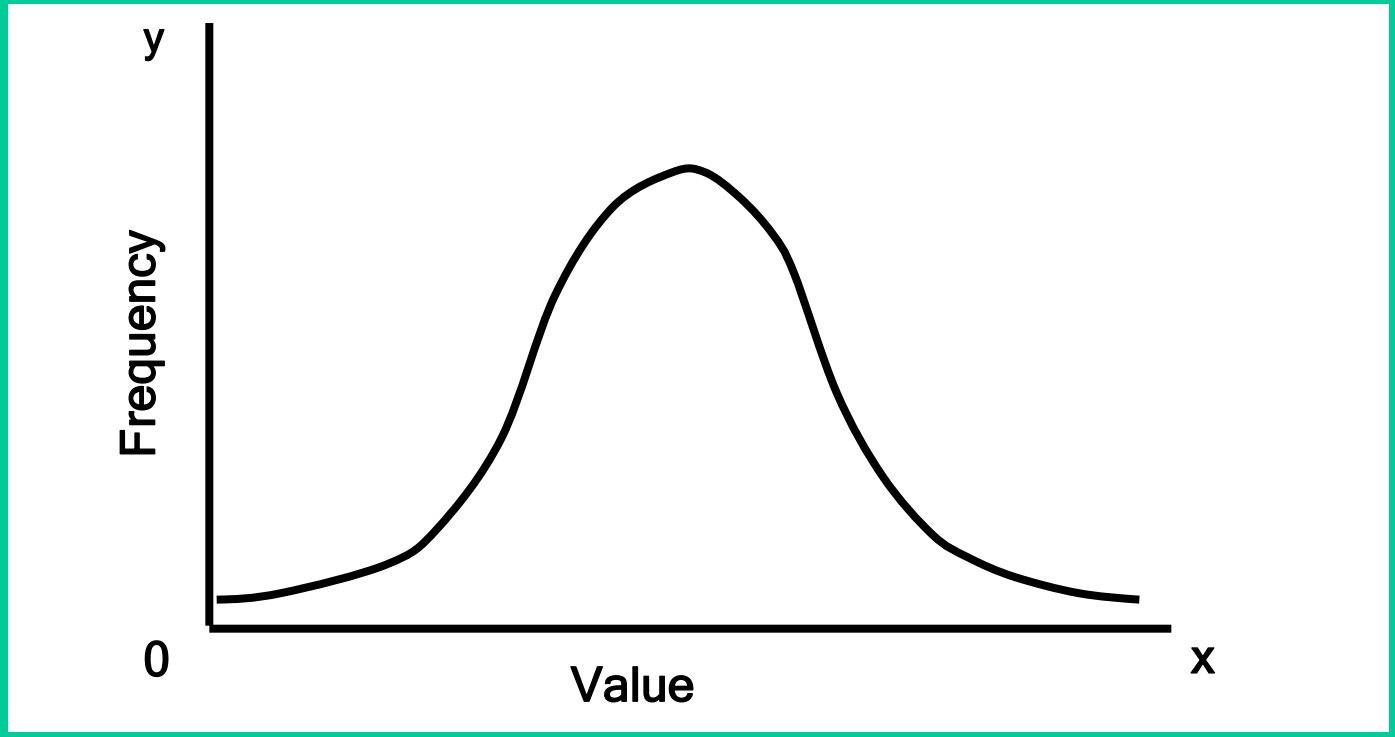
Example Calculation for Range

- Range: high score – low score = range

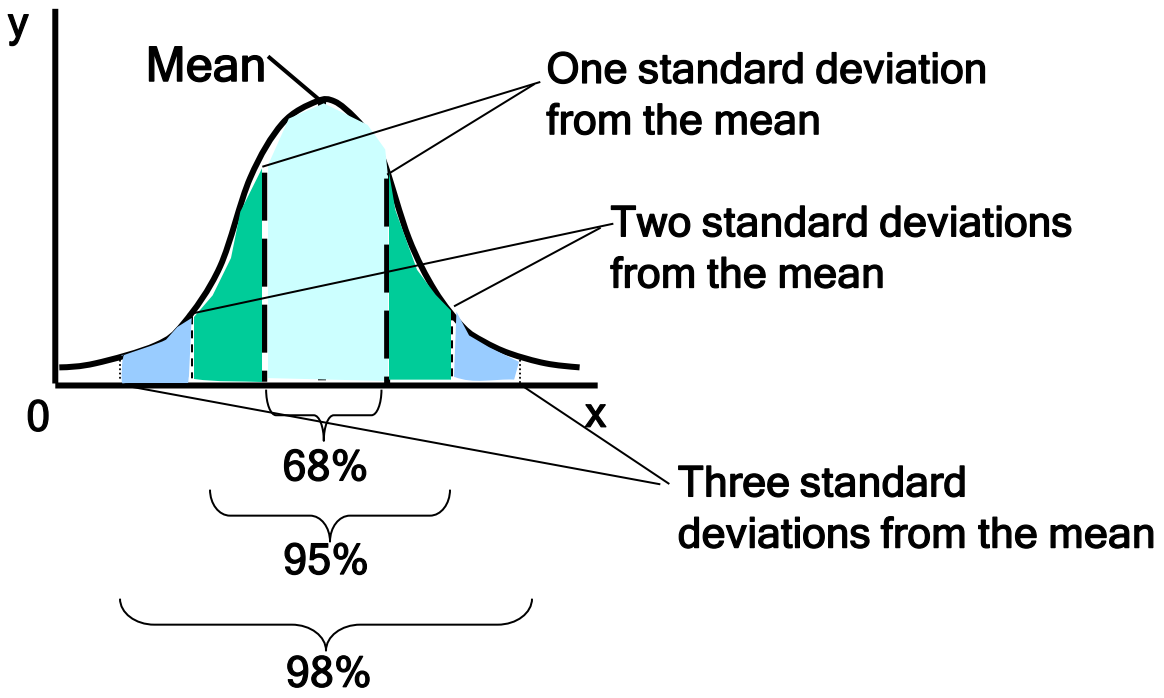
  range = 75 - 41

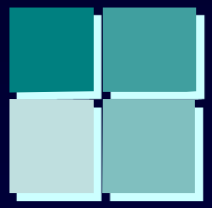
  range = 34

Normal Curve (Bell)



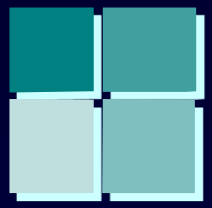
Standard Deviation





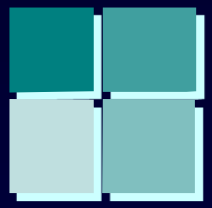
Calculating Standard Deviation

- Calculating is time consuming if have large N
- Can use statistical programs:
  - SPSS
  - Excel or other spreadsheet program



Guidelines for Analyzing Quantitative Survey Results

1. Choose a standard way to analyze the data and apply it consistently
2. Do not combine the middle category with each side of the scale
3. Do not report an “agree” or “disagree” category without also reporting the “strongly agree” or “strongly disagree” category
4. Analyze and report percentages and numbers
5. Provide the number of respondents a point of reference
6. If there is little difference in the data, raise the benchmark
7. Remember that data analysis is an art and a skill, it gets easier with training and practice



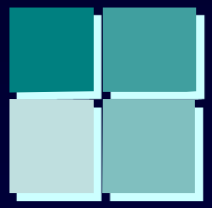
Describing Two Variables at the Same Time

- Two variables at once
- Example: What percent were boys and what percent were girls injured by UXO and by mines?

Example Two Variables at the Same Time

	Mines		UXO	
Boys (<16) N=62	28	55%	34	45%
Adult men N=63	22	45%	41	55%
Total 125	N=50	100%	N=75	100%

Source: Fabricated Data: 2004 Survey



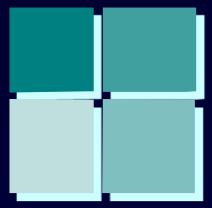
Two Variables with Crosstabs

- Cross tabulation (crosstab)
  - usually presented in a matrix format
  - displays two or more variables simultaneously
  - each cell shows number of respondents

Example Crosstabs

	Mines	UXO	Total %
Boys (<16) (n=62)	45%	55%	100%
Adult men (n=63)	35%	65%	100%

N=125 Source: Fabricated Data 2009



Variables

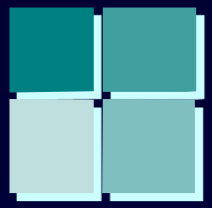
- Independent
  - Variable which you believe explains a change in the dependent variable
  - Program evaluation: the program
- Dependent
  - Variable you want to explain
  - Program evaluation: the outcomes

Example: Comparison of Means

-dependent variable: annual income

-independent variable: gender

	Mean Annual Income
Women N=854	27,800 SA Rand
Men N=824	32,400 SA Rand
Source: Fabricated data, 2009 survey	



Measure of Association

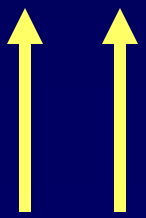
- How strongly variables are related, reported differently
- Measures of association (or relationship)
  - range from -1 to 1

Interpretation of Association

- Perfect relationship: 1 or -1
  - closer to 1 or -1: strong relationship
  - .5 moderate/strong (maybe as good as it gets)
- Closer to zero: no relationship
  - .2 slight/weak relationship

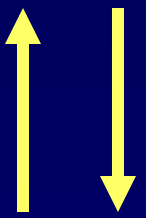
Direct Relationship

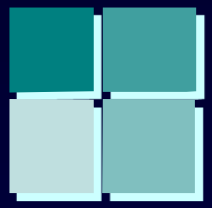
- Plus sign +
  - both variables change in the same direction
  - example:
    - as driving speed increases, death rate goes up



Inverse Relationship

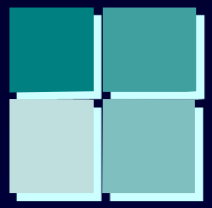
- Minus sign -
  - both variable change but in the opposite direction
  - example:
    - as age increases, health status decreases





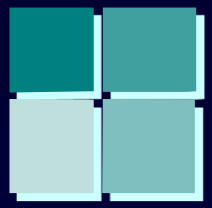
Inferential Statistics

- Used to analyze data from randomly selected samples
- Risk of error because your sample may be different from the population as a whole
- To make an inference, you first need to estimate the probability of that error



Statistical Significance Tests

- Tools to estimate how likely the results are in error
- Called tests of statistical significance
  - to estimate how likely it is that you have gotten the results you see in you analysis *by chance* alone



Statistical Significance

- Benchmark of .5%
  - .05 Alpha level or p-value
- Indicates:
  - we are 95% certain that our sample results are not due to chance

    or
  - the results are statistically significant at the .05 level
- Most reports do not go beyond .5

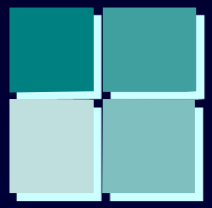
Chi Square and t-Test

Chi Square

- Not the strongest, but one of the most popular statistics
  - easy to calculate and interpret
- Used to compare two sets of nominal data (i.e. marital status and religious affiliation)
- Used to compare two ordinal variables or a combination of nominal and ordinal variables

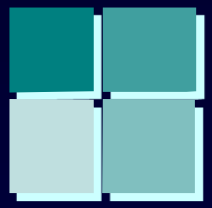
t-Test

- Used to determine if one group of numerical scores is statistically higher or lower than another group of scores
- Compares means for the groups
- Cumbersome for more than three groups



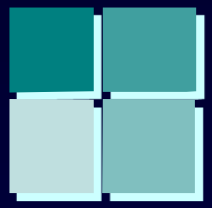
Analysis of Variance (ANOVA)

- Use to assess how nominal independent variables influence a continuous dependent variable
- Better than t-test for more than three groups
- Assumes populations have equal standard deviations and samples are randomly selected



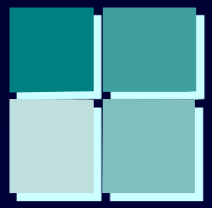
Remember:

- A significant test is nothing more than an estimate of the probability of getting the results by chance if there really is no difference in the population



Linking Qualitative and Quantitative Data

- Should qualitative and quantitative data and associated methods be linked during study design?
  - How?
  - Why?



Qualitative-Quantitative Linkages

- Confirmation or corroboration - triangulation
- Richer detail
- Initiate new lines of thinking
- Expand the scope

Completing the Design Matrix

- See example in Appendix 2 of the text

A Final Note….

"More fundamentally, students should be taught that instead of asking 'What techniques shall I use here?,' they should ask 'How can I summarize and understand the main features of this set of data?"

-- Chris Chatfield

Questions?