Briefing Document: CSC 439 - Statistical Processing Systems (Lecture 1)

Overview:

This document summarises key concepts presented in the first lecture of the "CSC 439 - Statistical Processing Systems" course, taught by Professor Benjamin Aribisala at Lagos State University. The lecture introduces fundamental principles of research design and data collection, emphasizing quantitative methods and the importance of sampling. It covers a range of topics, from survey methodologies to content analysis and structured observation.

Key Themes and Ideas:

1. **Course Outline and Assessment:**

- The course covers a wide range of statistical topics including design and analysis of sampling surveys, statistical data compression, frequency tables, estimation, hypothesis testing, analysis of categorical data, model validation, chi-square tests, and probability plots.
- Assessment is based on a combination of coursework (30%) consisting of projects, assignments, and tests, and a final examination (70%).
- Practical sessions will involve the use of Google Forms, Python, or R.

1. **The Research Process:**

- The lecture outlines a simplified view of the research process, progressing from research questions to research design, data collection/statistical analysis, and finally, findings.
- The process is not always linear: "Is research design a step on the way from research questions to findings? (Not necessarily…)"

1. **Quantitative Data Collection Methods:**

- The lecture describes various methods for quantitative data collection, including:
- Experiments/quasi-experiments ("'Controlled' comparisons Experiments / quasi-experiments (not common in some disciplines)")
- Surveys ("'Talking' with people Surveys")
- Structured observation ("Observing Structured observation")
- Secondary analysis of existing survey data ("Using existing quantitative information: I Secondary analysis of existing 'survey' (micro-)data")

- Analysis of existing published quantifiable data ("Using existing quantitative information: II Analysis of existing published quantifiable data")
- Content analysis ("Looking at 'texts' (books, films, web pages, adverts…) (Some forms of) content analysis")

1. **Experiments vs. Observational Studies:**

- Experiments allow for stronger causal inferences by randomly assigning participants to treatment and control groups. However, ethical and practical constraints often limit the feasibility of experiments in social research.
- Observational studies are more common but susceptible to "selection effects," where observed relationships may be due to pre-existing differences between groups rather than the variable of interest.
- Multivariate analysis is important to control for confounding factors: "It is thus important to control for other factors that may have induced an observed relationship. Controlling for such factors inevitably involves the use of a multivariate analysis…"

1. **Unobtrusive Methods and Subjectivity:**

- Some approaches, such as analyzing existing data, are "unobtrusive" and avoid the interactions inherent in surveys. However, all quantitative methods involve choices about "operationalisation of concepts and coding of data," which introduce subjectivity.

1. **Samples and Populations:**

- Statistical analyses are typically conducted on samples, but the ultimate goal is to make inferences about the larger population.
- Therefore, techniques are used to infer from the sample to the corresponding population.

1. **Sampling Error vs. Bias:**

- Sampling error is an inevitable consequence of studying a sample rather than the entire population.
- Bias ("non-sampling error") arises from deficiencies in the sampling approach.
- Researchers aim to estimate population parameters using sample observations, acknowledging that these estimates will include sampling error and potential bias. "So the parameter estimates generated by quantitative

research are equal to the population parameters, plus a certain amount of sampling error, plus any bias arising from the data 'collection' process."

1. **Types of Studies: Cross-sectional vs. Longitudinal:**

- Cross-sectional studies collect data at a single point in time. They are common but limited in their ability to examine social processes.
- Longitudinal studies observe the same population over an extended period, enabling analysis of change and stronger causal inferences. Types of longitudinal studies include:
- Trend studies: examine change within a population over time.
- Cohort studies: examine specific subpopulations over time.
- Panel studies: examine the same set of people each time.

1. **Survey Design Issues:**

- Key considerations in survey design include:
- Form of implementation (self-administered, interview, telephone).
- Defining the case, population, and sample.
- Choosing between probability and non-probability sampling methods.
- Determining the appropriate sample size.

1. **Survey Methods:**

- The lecture identifies three types of surveys: self-administered questionnaires, interview surveys, and telephone surveys.
- Each method has associated advantages and disadvantages, such as cost, response rate, interviewer bias, and the complexity of questions that can be asked.
- Self-completion is cheap and covers a wide area but has a low response rate.
- Telephone surveys are relatively cheap and quick, but people may not have home phones or may break off the survey easily.
- Face-to-face interviews have high response rates and enable clarification but are slow and expensive.

1. **Response Rate:**

- The response rate, calculated as the proportion of people selected who actually participate, is crucial in assessing the representativeness of the sample.

1. **Strengths and Weaknesses of Survey Research:**

- **Strengths:** Useful for describing large populations, feasible with large samples, flexible, reliable, and transparent.
- **Weaknesses:** Seldom deals with context, inflexible, subject to artificiality, potentially weak in validity, and may be poor at gathering certain sorts of information.

1. **Defining the Research Focus: Case, Population, and Sample:**

- **Case:** Each empirical instance of what is being researched.
- **Population:** All theoretically relevant cases.
- **Sample:** A selection of cases from the population.
- Sampling is necessary when the population is too large or costly to study in its entirety.

1. **Probability vs. Non-Probability Sampling:**

- **Probability Samples:** Allow for inferences about the population as a whole, based on sample results. The probability of selection is known.
- **Non-Probability Samples:** Do not formally allow for inferences about the population, but are used for pragmatic reasons. Inferential statistics are often (though controversially) generated from them.

1. **Types of Non-Probability Sampling:**

- **Reliance on available subjects:** Choosing people because they are available, justified only if less problematic methods are not possible.
- **Purposive or judgmental sampling:** Selecting a sample based on knowledge of the population and the study's purpose, especially useful for studying deviance.
- **Snowball sampling:** Using existing participants to locate other members of the population, suitable when the population is hard to reach.
- **Quota sampling:** Constructing a matrix of the population and collecting data from people matching the characteristics of each cell.

1. **Logic of Probability Sampling:**

- **Representativeness:** A sample is representative if it has the same aggregate characteristics as the population.

- **EPSEM (Equal Probability of Selection Method):** Every member of the population has the same chance of being selected.
- **Random sampling:** Each element has a known, non-zero chance of selection.
- **Sampling frame:** A list of every element in the population from which a sample can be selected.

1. **Types of Probability Sampling:**

- **Simple Random Sample:** Feasible only with a comprehensive sampling frame.
- **Systematic Random Sample:** Uses a random starting point, with every kth element selected.
- **Stratified sampling:** Selecting cases from homogeneous subsets of the population.
- **Multi-stage Sampling:** Involves repeating steps of creating lists and sampling from them.
- **Probability Proportional to Size (PPS) sampling:** A sophisticated form of multi-stage sampling where sampling units are selected with a probability proportional to their size.

1. **Sample Size Considerations:**

- Sample size depends on:
- The heterogeneity of the population.
- The number of relevant sub-groups.
- The frequency of the phenomenon being studied.
- The desired accuracy.
- The desired confidence level.
- Also consider the expected response rate and the planned forms of statistical analysis. "Generally (given a choice): Bigger is better! (hence the sample size often reflects costs/resources.)"

1. **Content Analysis:**

- A method of transforming qualitative data into a quantitative form by systematically coding the symbolic content of documents.

1. **Comparative-Historical Research:**

- Can involve statistical analysis, particularly when looking at change over time or comparing different countries.
- Uses macro-level secondary statistics (e.g., World Bank data) or primary statistics constructed by the researcher.
- Sampling considerations apply when analyzing a subset of historical events, with inferential statistics only appropriate for random samples.

1. **Observational Studies:**

- Quantitative methods can be applied to structured or systematic observation.
- Requires a pre-determined observation schedule or coding scheme, ensuring categories are mutually exclusive and exhaustive.
- Sampling considerations include the unit of analysis (events, interactions, or individuals) and the dimension of time.

1. **Benefits and Drawbacks of Structured Observation:**

- **Benefits:** Unobtrusive, avoids negotiation of meaning, produces reliable data, and enables large-scale data collection.
- **Drawbacks:** Limited to pre-determined categories, dependent on the researcher's ability to assess behaviour, and ahistorical.

Implications:

This lecture provides a foundational understanding of research design and quantitative methods. It underscores the importance of careful planning, appropriate sampling techniques, and awareness of potential biases in order to draw valid conclusions from data. The discussion of different data collection methods equips students with the knowledge to select the most suitable approach for their research questions, while the emphasis on sample size and response rates highlights the practical challenges and resource constraints often encountered in research.

This briefing doc should provide a comprehensive overview of the key concepts covered in the lecture excerpt.