Prof. Nguyen Khanh Doanh (Phenikaa University, Vietnam)

TEACHING PLAN

I. COURSE OVERVIEW

Title: *Introduction to Statistics*

Full description: This course provides an introductory overview of statistical methods applied in language and culture research, focusing on data collection, management, descriptive and inferential statistical techniques, and regression modeling. The course is designed for Master's and Ph.D. students in language and cultural studies, equipping them with practical data analysis skills using Stata (or R). By the end of the course, students will be able to independently conduct statistical analyses and apply statistical reasoning to language and cultural studies research.

Duration in hours: 18

Credits: 4 credits for Ph.D. students, 3 credits for MA students

Teaching language: English

II. COURSE STRUCTURE

This course is designed to provide students with a foundational understanding of statistical methods used in language and cultural research. The course is structured into six sessions, each focusing on a key aspect of statistical analysis:

- Session 1: Descriptive statistics for language and culture research
- Session 2: Introduction to statistical software (Stata/R)
- Session 3: Sampling methods and data collection in cultural studies
- Session 4: Data cleaning and preparation for language and culture research
- Session 5: Introduction to inferential statistics - hypothesis testing
- Session 6: Correlation and regression in language and culture studies

Each session includes theoretical explanations (40%) and hands-on practice (60%), ensuring students gain both conceptual knowledge and practical skills in statistical analysis.

<p align="center">**SESSION 1: DESCRIPTIVE STATISTICS FOR LANGUAGE AND CULTURE RESEARCH**</p>

Topic: Descriptive Statistics - Applications in Language and Culture Studies

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (R can be introduced if needed)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand key concepts of descriptive statistics and their role in research.
- Differentiate types of data (categorical vs. continuous, nominal vs. ordinal, etc.).
- Summarize data using measures of central tendency (mean, median, mode).
- Analyze data dispersion using variance, standard deviation, IQR, range.
- Visualize data using histograms, boxplots, bar charts, scatter plots in Stata.
- Apply descriptive statistics to real-world research in language and cultural studies.

2. DETAILED CONTENT (4 sessions, 50 minutes each)

Period 1: Descriptive Statistics, Types of Data, and Central Tendency (1 hour)

1.1. Overview of descriptive statistics

- What is descriptive statistics?
- Why is it important in language and culture studies?
- Real-life examples: Analyzing language fluency across cultures, measuring cultural attitudes using survey data.

1.2. Types of data and their implications

- Categorical vs. numerical data.
- Nominal, ordinal, interval, ratio scales.
- How does data type affect statistical analysis?
- Case Study: Data classification in linguistic diversity research (e.g., language family classification).

1.3. Measures of central tendency

- Mean, Median, Mode - What do they tell us about data?
- When to use mean vs. median (importance in skewed data).
- Case study: Language fluency scores across different cultural groups.

- Hands-on activity (Stata practice)

Period 2: Measures of Dispersion and Data Visualization (1 hour)

2.1. Measures of dispersion and variability

- Range, interquartile Range (IQR), variance, standard deviation.
- Why measuring variability is crucial in cross-cultural analysis.
- Example: Analyzing variation in dialect usage across regions.

2.2. Data visualization

- The Power of data visualization in language and culture research.
- Common graph types and when to use them:
- Histograms - Distribution of numerical data.
- Boxplots - Identifying outliers and spread.
- Bar Charts - Categorical data comparison.
- Scatter Plots - Exploring relationships between two variables.
- Tips for choosing colors and formats suitable for academic publishing.
- Hands-on activity (Stata practice)

Period 3: Applying Descriptive Statistics in Research & Reporting Results (1 hour)

3.1. Reporting descriptive statistics in academic writing

- APA style for presenting summary statistics.
- How to interpret and explain results in a research paper.
- Example: Writing a results section for a study on cultural attitudes.
- Common mistakes in reporting descriptive results.

3.2. Applying descriptive statistics to student research

- Students apply descriptive statistics techniques to their own datasets.
- Group discussion: Challenges and solutions in statistical analysis.
- Q&A session & research consultation on individual research projects.
- Hands-on activity (Stata practice)

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students analyze a provided dataset (e.g., language use survey) and submit a 1-page report summarizing their findings using descriptive statistics and visualizations.
- Provide detailed feedback on their reports, focusing on statistical accuracy, clarity, and APA formatting.

Feedback

- Provide detailed feedback on students' statistical calculations, interpretation of summary measures, and visualization choices.
- Assess accuracy in computing mean, median, mode, variance, standard deviation, and IQR.
- Evaluate clarity in explaining data insights, particularly when comparing linguistic and cultural datasets.
- Highlight common mistakes, such as misuse of mean for skewed data, incorrect selection of visualizations, or failure to justify findings.
- Offer suggestions for improvement, including best practices in APA formatting, effective graph labeling, and refining data narratives for research papers.

5. ADDITIONAL RESOURCES

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2020). *Statistical Techniques in Business and Economics* (18 ed.). USA: McGraw Hill.

SESSION 2: INTRODUCTION TO STATISTICAL SOFTWARE (STATA/R)

Topic: Getting Started with Stata/R for Language and Culture Research

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (R can be introduced if needed)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand the basic interface and functionality of Stata (or R).
- Import, manage, and explore datasets in Stata/R.
- Perform basic data manipulation tasks (e.g., sorting, filtering, creating new variables).
- Generate summary statistics and simple visualizations using Stata/R.
- Apply Stata/R skills to real-world research in language and cultural studies.

2. DETAILED CONTENT (3 periods, 1 hour each)

Period 1: Introduction to Stata/R and Data Management (1 hour)

1.1. Overview of Stata/R

- What is Stata/R, and why is it useful for research?
- Key features and advantages of Stata/R in language and culture studies.
- Real-life examples.

1.2. Getting started with Stata/R

- Installing and setting up Stata/R.
- Understanding the interface: Command window, Data Editor, Results window.
- Basic commands and syntax in Stata/R.

1.3. Data management basics

- Importing data from Excel, CSV, or other formats.
- Exploring datasets: Viewing data, checking variable types, and identifying missing values.
- Saving and exporting data.
- Hands-on activity (Stata/R practice)

Period 2: Data Manipulation and Summary Statistics (1 hour)

2.1. Data manipulation in Stata/R

- Sorting and filtering data.
- Creating new variables (e.g., recoding, calculating means).
- Handling missing data.

2.2. Generating summary statistics

- Calculating measures of central tendency (mean, median, mode).
- Calculating measures of dispersion (variance, standard deviation, range).
- Example: Summarizing language fluency scores across cultural groups.
- Hands-on activity (Stata/R practice).

Period 3: Data Visualization and Applying Skills to Research (1 hour)

3.1. Data visualization in Stata/R

- Introduction to basic graphs: Histograms, bar charts, scatter plots.
- Customizing graphs: Titles, labels, colors.
- Example: Visualizing language diversity across regions.

3.2. Applying Stata/R skills to research

- How to use Stata/R for real-world research in language and culture studies.
- Case study: Analyzing and visualizing data from a cultural attitudes survey.
- Hands-on activity (Stata/R practice).

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students import a provided dataset (e.g., cultural attitudes survey) into Stata/R.
- Perform basic data manipulation tasks (e.g., create a new variable, filter data).
- Generate summary statistics and one visualization (e.g., histogram or bar chart).
- Submit a 1-page report summarizing their findings and the steps they took in Stata/R.

Feedback

- Provide detailed feedback on their Stata/R commands, results, and visualizations.
- Highlight areas for improvement in data management and analysis.

4. ADDITIONAL RESOURCES

Acock, A. C. (2018). *A Gentle Introduction to Stata* (6th ed.). Stata Press.

SESSION 3: SAMPLING METHODS AND DATA COLLECTION IN CULTURAL STUDIES

Topic: Designing Sampling Strategies and Collecting Data for Language and Culture Research

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (optional for data exploration)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand the importance of sampling methods in research design.
- Differentiate between various sampling techniques (e.g., random, stratified, cluster).
- Design a sampling strategy for a research project in language and culture studies.
- Identify and address common challenges in data collection for cultural research.
- Apply sampling and data collection techniques to real-world research scenarios.

2. DETAILED CONTENT (3 periods, 1 hour each)

Period 1: Introduction to Sampling Methods (1 hour)

1.1. Overview of sampling in research

- What is sampling, and why is it important?
- Key concepts: Population, sample, sampling frame, and representativeness.
- Real-life examples.

1.2. Types of sampling methods

- Probability sampling
- Non-probability sampling

1.3. Choosing the right sampling method

- Factors to consider: Research objectives, population characteristics, and resources.
- Case study: Selecting a sampling method for a study on language attitudes in a multicultural city.
- Hands-on activity: Students work in groups to design a sampling strategy for a hypothetical research project (e.g., studying dialect variations in rural vs. urban areas).

Period 2: Data Collection Techniques and Challenges (1 hour)

2.1. Data collection methods

- Surveys and questionnaires: Designing effective questions for cultural research.
- Interviews: Structured vs. semi-structured approaches.
- Ethnographic methods: Participant observation and field notes.
- Example: Collecting data on language use in a bilingual community.

2.2. Challenges in data collection

- Language barriers and translation issues.
- Cultural sensitivity and ethical considerations.
- Ensuring data quality and reliability.
- Case study: Addressing challenges in a study on cultural attitudes toward language preservation.
- Hands-on activity: Students design a data collection tool (e.g., a survey or interview guide) for their hypothetical research project.

Period 3: Applying Sampling and Data Collection Skills to Research (1 hour)

3.1. Integrating sampling and data collection into research design

- Aligning sampling methods with research questions.
- Ensuring ethical compliance in data collection.
- Example: Designing a study on the impact of globalization on indigenous languages.

3.2. Practical application

- Students apply their sampling and data collection skills to a real-world scenario.
- Case study: Analyzing a dataset collected from a cultural attitudes survey.
- Hands-on activity (optional Stata practice).

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students design a sampling strategy and data collection plan for their own research project (or a hypothetical project).
- Submit a 1-page report outlining:
- Research question and objectives.
- Sampling method and justification.
- Data collection tools and procedures.
- Potential challenges and solutions.

Feedback

- Provide detailed feedback on their sampling and data collection plans.
- Highlight strengths and suggest improvements in methodology and practical implementation.

4. ADDITIONAL RESOURCES

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2020). *Statistical Techniques in Business and Economics* (18 ed.). USA: McGraw Hill.

SESSION 4: DATA CLEANING AND PREPARATION FOR LANGUAGE AND CULTURE RESEARCH

Topic: Preparing and Cleaning Data for Analysis in Language and Culture Studies

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (R can be introduced if needed)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand the importance of data cleaning and preparation in research.
- Identify and address common data issues (e.g., missing data, outliers, inconsistencies).
- Perform basic data cleaning tasks using Stata (or R).
- Prepare datasets for statistical analysis in language and culture studies.
- Apply data cleaning and preparation techniques to real-world research scenarios.

2. DETAILED CONTENT (3 periods, 1 hour each)

Period 1: Introduction to Data Cleaning and Preparation (1 hour)

1.1. Overview of data cleaning and preparation

- What is data cleaning, and why is it important?
- Key concepts: Data quality, consistency, and reliability.
- Real-life examples.

1.2. Common data issues in language and culture research

- Missing data: Causes and solutions.
- Outliers: Identifying and handling extreme values.
- Inconsistencies: Typos, formatting errors, and duplicate entries.
- Case study: Cleaning a dataset on cultural attitudes toward language preservation.

1.3. Data preparation workflow

- Steps in data cleaning: Importing, exploring, cleaning, and exporting.
- Tools and software for data cleaning (Stata, R, Excel).
- Hands-on activity (Stata practice).

Period 2: Data Cleaning Techniques (1 hour)

2.1. Handling missing data

- Types of missing data: MCAR, MAR, MNAR.
- Strategies for dealing with missing data: Deletion, imputation, and analysis.
- Example: Imputing missing values in a dataset on language fluency.

2.2. Identifying and addressing outliers

- Methods for detecting outliers: Boxplots, z-scores.
- Deciding whether to remove or transform outliers.
- Example: Analyzing outliers in a dataset on cultural attitudes.

2.3. Fixing inconsistencies and errors

- Standardizing variable names and formats.
- Removing duplicate entries.
- Correcting typos and formatting errors.
- Hands-on activity (Stata practice).

Period 3: Preparing Data for Analysis and Applying Skills to Research (1 hour)

3.1. Preparing datasets for analysis

- Creating new variables (e.g., recoding, calculating means).
- Merging and reshaping datasets.
- Example: Preparing a dataset for regression analysis on language use.

3.2. Applying data cleaning and preparation skills to research

- Students apply techniques to their own datasets (or a provided dataset).
- Case study: Cleaning and preparing a dataset on dialect variations.
- Hands-on activity (Stata practice).

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students clean and prepare a provided dataset (e.g., cultural attitudes survey).
- Submit a 1-page report outlining:
- Data issues identified and how they were addressed.
- Steps taken to clean and prepare the dataset.
- Challenges encountered and solutions implemented.

Feedback

- Provide detailed feedback on their data cleaning and preparation process.
- Highlight strengths and suggest improvements in methodology and practical implementation.

4. ADDITIONAL RESOURCES

Long, J. S. (2008). *The Workflow of Data Analysis Using Stata*. Stata Press.

SESSION 5: INTRODUCTION TO INFERENTIAL STATISTICS: HYPOTHESIS TESTING

Topic: Understanding and Applying Hypothesis Testing in Language and Culture Research

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (R can be introduced if needed)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand the basic concepts of inferential statistics and hypothesis testing.
- Formulate null and alternative hypotheses for research questions.
- Perform and interpret common hypothesis tests (e.g., t-test, chi-square test).
- Apply hypothesis testing techniques to real-world research in language and culture studies.
- Report hypothesis testing results in academic writing.

2. DETAILED CONTENT (3 periods, 1 hour each)

Period 1: Introduction to Inferential Statistics and Hypothesis Testing (1 hour)

1.1. Overview of inferential statistics

- What is inferential statistics, and why is it important?
- Key concepts: Population, sample, sampling distribution, and statistical significance.
- Real-life examples.

1.2. Hypothesis testing basics

- Null hypothesis (H_0) vs. alternative hypothesis (H_1).
- Types of errors: Type I and Type II errors.
- Significance level (α) and p-values.
- Case study: Formulating hypotheses for a study on dialect variations.

1.3. Steps in hypothesis testing

- Step 1: State the hypotheses.
- Step 2: Choose the appropriate test.
- Step 3: Calculate the test statistic.

- Step 4: Make a decision (reject or fail to reject H$_0$).
- Step 5: Interpret the results.
- Hands-on activity (Stata practice).

Period 2: Common Hypothesis Tests (1 hour)

2.1. T-tests

- Independent samples t-test: Comparing means between two groups.
- Paired samples t-test: Comparing means within the same group over time.
- Example: Testing differences in language fluency scores between two cultural groups.

2.2. Chi-square tests

- Chi-square test of independence: Testing relationships between categorical variables.
- Example: Analyzing the association between language use and cultural identity.

2.3. Choosing the right test

- Factors to consider: Type of data, research question, and assumptions.
- Case study: Selecting the appropriate test for a study on cultural attitudes.
- Hands-on activity (Stata practice).

Period 3: Applying Hypothesis Testing to Research and Reporting Results (1 hour)

3.1. Applying hypothesis testing to research

- Students apply hypothesis testing techniques to their own datasets (or a provided dataset).
- Case study: Testing hypotheses about the impact of globalization on indigenous languages.

3.2. Reporting hypothesis testing results

- APA style for reporting statistical results.
- How to interpret and explain p-values, test statistics, and confidence intervals.
- Common mistakes in reporting hypothesis testing results.
- Hands-on activity (Stata practice).

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students perform hypothesis tests on a provided dataset (e.g., cultural attitudes survey).
- Submit a 1-page report outlining:
- Research question and hypotheses.
- Test performed and justification.
- Results and interpretation.

- Implications for language and culture research.

Feedback

- Provide detailed feedback on their hypothesis testing process and results.
- Highlight strengths and suggest improvements in methodology and reporting.

4. ADDITIONAL RESOURCES

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2020). *Statistical Techniques in Business and Economics* (18 ed.). USA: McGraw Hill.

SESSION 6: CORRELATION AND REGRESSION IN LANGUAGE AND CULTURE STUDIES

Topic: Exploring Relationships Between Variables in Language and Culture Research

Target audience: Master's & Ph.D. students - Department of Studies on Language and Culture

Duration: 3 hours

Teaching approach: 40% theory - 60% hands-on practice

Software used: Stata (R can be introduced if needed)

1. COURSE OBJECTIVES

By the end of the course, students will be able to:

- Understand the concepts of correlation and regression analysis.
- Calculate and interpret correlation coefficients.
- Perform and interpret simple and multiple regression analyses.
- Apply correlation and regression techniques to real-world research in language and culture studies.
- Report correlation and regression results in academic writing.

2. DETAILED CONTENT (3 periods, 1 hour each)

Period 1: Introduction to Correlation Analysis (1 hour)

1.1. Overview of correlation analysis

- What is correlation, and why is it important?
- Key concepts: Positive, negative, and zero correlation.
- Real-life examples.

1.2. Types of correlation coefficients

- Pearson's r: For linear relationships between continuous variables.
- Spearman's rho: For monotonic relationships (ordinal or non-normal data).
- Case study: Calculating correlation coefficients for a dataset on language attitudes.

1.3. Interpreting correlation results

- Strength and direction of correlation.
- Statistical significance and confidence intervals.
- Example: Interpreting the correlation between language proficiency and cultural integration.
- Hands-on activity (Stata practice).

Period 2: Introduction to Regression Analysis (1 hour)

2.1. Overview of regression analysis

- What is regression, and why is it useful?
- Key concepts: Dependent vs. independent variables, regression coefficients.
- Real-life examples.

2.2. Simple linear regression

- Model formulation: $Y = a + bX + u$
- Interpreting regression coefficients (a and b).
- Assessing model fit: R-squared and residuals.
- Example: Predicting language proficiency from cultural exposure.

2.3. Multiple linear regression

- Model formulation: $Y = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n + u$
- Interpreting multiple regression coefficients.
- Case study: Analyzing the impact of multiple factors (e.g., age, education, cultural exposure) on language use.
- Hands-on activity (Stata practice).

Period 3: Applying Correlation and Regression to Research and Reporting Results (1 hour)

3.1. Applying correlation and regression to research

- Students apply correlation and regression techniques to their own datasets (or a provided dataset).
- Case study: Exploring relationships between language use and cultural attitudes.

3.2. Reporting correlation and regression results

- APA style for reporting correlation and regression results.
- How to interpret and explain regression coefficients, R-squared, and p-values.
- Common mistakes in reporting correlation and regression results.
- Hands-on activity (Stata practice).

3. ASSESSMENT AND FEEDBACK

Short assignment

- Students perform correlation and regression analyses on a provided dataset (e.g., cultural attitudes survey).

- Submit a 1-page report outlining:
- Research question and hypotheses.
- Correlation and regression results.
- Interpretation and implications for language and culture research.

Feedback

- Provide detailed feedback on their correlation and regression analyses.
- Highlight strengths and suggest improvements in methodology and reporting.

4. ADDITIONAL RESOURCES

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2020). *Statistical Techniques in Business and Economics* (18 ed.). USA: McGraw Hill.