

Content

First Half Topics

Introduction: Basic Concepts

- Data
- Cases
- Value
- Unit of Analysis
- Estimator
- Two main goals of data analysis
- Hypotheses
- Dependent Variable
- Independent Variable
- Population
- Sample
- Parameters
- Statistics
- Descriptive Statistics
- Inferential Statistics

Unit I: Descriptive Statistics

- Kinds of variables
- Measurement metrics
- “Exploratory data analysis”
- Kinds of univariate graphics (histogram, box plots, etc)
- Tabulations and crosstabulations
- Shape, center, spread, skew
- Mistakes with graphics: axes, scales, etc
- Measures of central tendency
- Measures of dispersion

- Outliers
- Linear transformations
- Density curves
- Normal distribution
- Z-scores
- Cumulative probabilities

Unit II: Statistical Relationships

- Kinds of bivariate graphics (scatterplots, linear fits, smooth lines, etc)
- Scatterplot diagnoses: form, direction, strength, outliers
- Transformed data in scatterplots
- Correlation coefficients
- Sample vs. population calculation differences
- Regression
 - intercept
 - slope
 - model component
 - stochastic component
 - sources of error in modeling
 - method of least squares
 - model sum of squares
 - residual/error sum of squares
 - total sum of squares
 - R^2
 - \hat{Y}
 - Y_i
 - \bar{Y}
 - Calculating the slope (b_1) from a correlation coefficient
 - Regression coefficients vs correlation coefficients
 - Residual plots
 - Effects of outliers on regression lines
 - Predicted values/expected values of Y given X
- Two-way tables for qualitative variables

Unit III: Producing Data & Research Design

- Correlation vs. causation
- Confounding variables
- Conditions for causation

- credible causal link (theory)
 - temporal precedence (could Y cause X?)
 - covariation (do X and Y move together/covary?)
 - no plausible alternative explanations (no confounding variables)
- Simpson's paradox
- Internal validity
- Threats to internal validity
 - History
 - Maturation/Learning
 - Testing
 - Instrumentation
 - Regression to the mean
 - Selection bias
 - Mortality
 - Social Desirability
- External validity
- Threats to external validity
 - context
 - sampling procedures
- Anecdotal data
- Survivor's bias
- Control through randomization
- Experiments
 - treatment
 - control
 - observations
 - random assignment
- Surveys
 - Population/sample/sampling frame
 - Simple random sample
 - Stratified random sample
 - Multi-stage sampling
 - Response rate
 - Response bias
 - Question framing and ordering effects
 - Bias vs. variability (validity vs reliability)

Unit IV: Probability

- Random phenomenon
- Independence
- Probability model
- Event
- Long run
- Rules of probability
 - Range
 - Sample space
 - Addition
 - Complementarity
 - Multiplication
- Disjoint events
- Random variable
 - Discrete
 - Continuous
- Probability distribution
- Expected value (mean) of a random variable
- Law of large numbers
- Standard deviation/variance of a random variable
- Rules for means of random variables
 - Addition
 - Subtraction
 - Linear transformations
- Conditional probability
- Probability trees

Second Half Topics

Unit V: Sampling

- Sample Distribution
- Population Distribution
- Sampling Distribution
- Central Limit Theorem
- Estimators as Random Variables
- Standard Error of the Mean
- Standard Error of a Proportion

- Standard Error of a Count

Unit VI: Inference

- Confidence Intervals
 - Point Estimate
 - Margin of Error
 - Critical Values
 - Margin of Error
 - Ways to reduce margin of error
 - Statistical Power
 - Inference when σ is unknown
 - Student's T distribution
 - Degrees of freedom (df)

Unit VII: Hypothesis Testing

- Logic of Hypothesis testing
- Null Hypothesis (H_0)
- Alternative Hypothesis (H_A)
- Types of Hypothesis Test
 - Tabular Analysis
 - Difference in Means
 - Correlation Coefficient
 - Regression
- How to choose test statistic
 - Critical Values (Z , T , χ^2)
 - One-Tail vs Two-Tail Tests
 - P-values
 - Type I and Type II Errors

Unit VIII: Simple Linear Regression

- Bivariate Regression
- Dichotomous Dummy Variables
- Difference in Means Tests via Regression
- Statistical Significance (hypothesis testing, confidence intervals, t- and p- values, etc)
- Substantive/Practical Significance (effect magnitudes, predicted values, R^2)

Unit IX: Less Simple Linear Regression

- Multiple Regression
- Confounding variables in Multiple Regression
- Multi-category Variables and Dummy Variables
- Standardized coefficients
- Simple model comparison: R^2 , effect magnitudes, predicted values

Stata Skills

- Be able to:
 - Load data
 - Use a .do file
 - Place comments in .do files
 - Calculate summary statistics
 - Graph scatterplots and histograms
 - Recode data
 - Tabulate and cross-tabulate data
 - Conduct difference in means tests
 - Run regressions
 - Predict values from regression results
- Selected Stata commands
`cd, use, clear, tab, sum, tabstat, gen, sort, browse, twoway, scatter, histogram, graph export, corr, reg, disp, bys, predict, save, keep, drop, if, recode, collapse, preserve, restore, ttest`