

Applied Statistics (Code:)

Syllabus

Physics Education Study Program Faculty of Mathematics and Science

Instructor : Dr. Dadan Rosana, M.Si.
Office : Phsics Education Study Program
Office hours :
Email : danrosana.uny@gmail.com

Required module:

Applied Statistics

Dadan Rosana, FMIPA UNY

Course Description:

This course serves two purposes. The first purpose of this course is to provide you with a background in statistical principles in order for you to be a good user of statistical analysis. We will learn how to describe data effectively, how to run a simple regression, and how to interpret the results. The second purpose of this course is to provide you with the basic knowledge in probability theories, such as expected values or probability distributions, which are necessary in understanding other courses in physics education research.

Learning Objectives

After careful study of this chapter you should be able to do the following:

1. Identify the role that statistics can play in the science education problem-solving process
2. Discuss how variability affects the data collected and used for making educational research decisions
3. Explain the difference between enumerative and analytical studies
4. Discuss the different methods that scientist use to collect data
5. Identify the advantages that designed experiments have in comparison to other methods of collecting science education data
6. Explain the differences between mechanistic models and empirical models
7. Discuss how probability and probability models are used in science education

Course Outline:

CHAPTER 1. The Role of Statistics in Educational Research

- 1-1 The Educational Research Method and Statistical Thinking
- 1-2 Collecting Education Research Data
 - 1-2.1 Basic Principles
 - 1-2.2 Retrospective Study
 - 1-2.3 Observational Study
 - 1-2.4 Designed Experiments

CHAPTER 2 Descriptive Statistics

- 2-1 Frequency Distribution
- 2-2 Distribution of Proportion
- 2-3 Population Parameters
- 2-4 Sample Statistics

CHAPTER 3 Probability

- 3-1 Sample Spaces and Events
 - 3-1.1 Random Experiments
 - 3-1.2 Sample Spaces
 - 3-1.3 Events
- 3-2 Counting Techniques
 - 3-2.1 Introduction
 - 3-2.2 Axioms of Probability
- 3-3 Addition Rules
- 3-4 Conditional Probability
- 3-5 Multiplication and Total Probability
 - 3-5.1 Multiplication Rule
 - 3-5.2 Total Probability Rule
- 3-6 Independence
- 3-7 Bayes' Theorem
- 3-8 Random Variables

CHAPTER 4 Probability Distributions 1 (Discrete Random Variables)

- 4-1 Discrete Random Variables
- 4-2 Probability Distributions and Probability Mass Functions
- 4-3 Cumulative Distribution Functions
- 4-4 Mean and Variance of a Discrete Random Variable
- 4-5 Discrete Uniform Distribution
- 4-6 Binomial Distribution
- 4-7 Geometric and Negative Binomial Distributions
 - 4-7.1 Geometric Distribution
 - 4-7.2 Negative Binomial Distribution
- 4-8 Hypergeometric Distribution
- 4-9 Poisson Distribution

CHAPTER 5 Probability Distributions 2 (Continuous Random Variables)

- 5-1 Continuous Random Variables
- 5-2 Probability Distributions and Probability Density Function

- 5-3 Cumulative Distribution Functions
- 5-4 Mean and Variance of a Continuous Random Variable
- 5-5 Continuous Uniform Distribution
- 5-6 Normal Distribution
- 5-7 Normal Approximation to the Binomial and Poisson Distributions
- 5-8 Exponential Distribution
- 5-9 Lognormal Distribution

CHAPTER 6 Random Sampling and Data Description

- 6-1 Data Summary and Display
- 6-2 Random Sampling
- 6-3 Stem-and-Leaf Diagrams
- 6-4 Frequency Distributions and Histograms
- 6-5 Box Plots
- 6-6 Time Sequence Plots
- 6-7 Probability Plots

CHAPTER 6 Point Estimation of Parameters

- 7-1 Introduction
- 7-2 General Concepts of Point Estimation
 - 7-2.1 Unbiased Estimators
 - 7-2.3 Variance of a Point Estimator
 - 7-2.4 Standard Error: Reporting a Point Estimator
 - 7-2.6 Mean Square Error of an Estimator
- 7-3 Methods of Point Estimation
 - 7-3.1 Method of Moments
 - 7-3.2 Method of Maximum Likelihood
 - 7-3.3 Bayesian Estimation of Parameters
- 7-4 Sampling Distributions 238
- 7-5 Sampling Distribution of Means

CHAPTER 7 Tests of Hypotheses for a Single Sample

- 8-1 Hypothesis Testing
 - 8-1.1 Statistical Hypotheses
 - 8-1.2 Tests of Statistical Hypotheses
 - 8-1.3 One-Sided and Two-Sided Hypotheses
 - 8-1.4 General Procedure for Hypothesis Testing
- 8-2 Tests on the Mean of a Normal Distribution, Variance Known
 - 8-2.1 Hypothesis Tests on the Mean
 - 8-2.2 P -Values in Hypothesis Tests
 - 8-2.3 Connection Between Hypothesis Tests and Confidence Intervals
 - 8-2.4 Type II Error and Choice of Sample Size
 - 8-2.5 Large Sample Test
 - 8-2.6 Some Practical Comments on Hypothesis Tests
- 8-3 Tests on the Mean of a Normal Distribution, Variance Unknown
 - 8-3.1 Hypothesis Tests on the Mean
 - 8-3.2 P -Value for a t -Test

- 8-3.3 Choice of Sample Size
- 8-4 Tests on the Variance and Standard Deviation of a Normal Distribution
 - 8-4.1 The Hypothesis Testing Procedures
 - 8-4.2 α -Error and Choice of Sample Size
- 8-5 Tests on a Population Proportion
 - 8-5.1 Large-Sample Tests on a Proportion
 - 8-5.2 Small-Sample Tests on a Proportion
 - 8-5.3 Type II Error and Choice of Sample Size
- 8-6 Summary of Inference Procedures for a Single Sample
- 8-7 Testing for Goodness of Fit
- 8-8 Contingency Table Tests

CHAPTER 9 Statistical Inference for Two Samples

- 9-1 Introduction
- 9-2 Inference For a Difference in Means of Two Normal Distributions, Variances Known
 - 9-2.1 Hypothesis Tests for a Difference in Means, Variances Known
 - 9-2.2 Choice of Sample Size
 - 9-2.3 Identifying Cause and Effect
 - 9-2.4 Confidence Interval on a Difference in Means, Variances Known
- 9-3 Inference For a Difference in Means of Two Normal Distributions, Variances Unknown
 - 9-3.1 Hypothesis Tests for a Difference in Means, Variances Unknown
 - 9-3.2 More About the Equal Variance Assumption
 - 9-3.3 Choice of Sample Size 344
 - 9-3.4 Confidence Interval on a Difference in Means, Variances Unknown
- 9-4 Paired t -Test
- 9-5 Inference on the Variances of Two Normal Distributions
 - 9-5.1 The F Distribution 355
 - 9-5.2 Development of the F Distribution
 - 9-5.3 Hypothesis Tests on the Ratio of Two Variances
 - 9-5.4 α -Error and Choice of Sample Size
 - 9-5.5 Confidence Interval on the Ratio of Two Variances
- 9-6 Inference on Two Population Proportions
 - 9-6.1 Large-Sample Test for $H_0 : p_1 = p_2$
 - 9-6.2 Small Sample Test for $H_0 : p_1 = p_2$
 - 9-6.3 Error and Choice of Sample Size
 - 9-6.4 Confidence Interval for $P_1 = P_2$
- 9-7 Summary Table for Inference Procedures for Two Samples

CHAPTER 10 Simple Linear Regression and Correlation

- 10-1 Empirical Models
- 10-2 Simple Linear Regression
- 10-3 Properties of the Least Squares Estimators
- 10-4 Some Comments on Uses of Regression
- 10-5 Hypothesis Tests in Simple Linear Regression

- 10-5.1 Use of t -Tests
- 10-5.2 Analysis of Variance Approach to Test Significance of Regression
- 10-6 Confidence Intervals
 - 10-6.1 Confidence Intervals on the Slope and Intercept
 - 10-6.2 Confidence Interval on the Mean Response
- 10-7 Prediction of New Observations
- 10-8 Adequacy of the Regression Model
 - 10-8.1 Residual Analysis
 - 10-8.2 Coefficient of Determination (R^2)
 - 10-8.3 Lack-of-Fit Test
- 10-9 Transformations to a Straight Line
- 10-10 More About Transformations
- 10-11 Correlation

CHAPTER 11 Multiple Linear Regression

- 11-1 Multiple Linear Regression Model
 - 11-1.1 Introduction
 - 11-1.2 Least Squares Estimation of the Parameters
 - 11-1.3 Matrix Approach to Multiple Linear Regression
 - 11-1.4 Properties of the Least Squares Estimators
- 11-2 Hypothesis Tests in Multiple Linear Regression
 - 11-2.1 Test for Significance of Regression
 - 11-2.2 Tests on Individual Regression Coefficients and Subsets of Coefficients
 - 11-2.3 More About the Extra Sum of Squares Method
- 11-3 Confidence Intervals in Multiple Linear Regression
 - 11-3.1 Confidence Intervals on Individual Regression Coefficients
 - 11-3.2 Confidence Interval on the Mean Response
- 11-4 Prediction of New Observations
- 11-5 Model Adequacy Checking
 - 11-5.1 Residual Analysis
 - 11-5.2 Influential Observations
- 11-6 Aspects of Multiple Regression Modeling
 - 11-6.1 Polynomial Regression Models
 - 11-6.2 Categorical Regressors and Indicator Variables
 - 11-6.3 Selection of Variables and Model Building
 - 11-6.4 Multicollinearity

Grading:

Homework	(30%)
Midterm exam	(30%): In-class, closed-book exam
Final exam	(40%): In-class, closed-book exam

More on Prerequisites

The most important prerequisite or background material for this course is outlined below. Several families of probability distributions are commonly used in regression modelling. For this course, the most important by far are the Normal (or Gaussian), t , F and χ^2 families, with the normal distribution being the most important. I expect you to be familiar with the properties of the (univariate) normal distribution as a prerequisite. I strongly recommend that you spend an hour or so on your own reading about the basic properties of the t , F , and χ^2 distributions, in particular how they are derived from normal samples, their parametrizations, and their means and variances. The text assumes you already know this. Any text on mathematical statistics will have this material, for example *Mathematical Statistics and Data Analysis*, 2nd Ed. by John A. Rice, Duxbury Press, 1994. The text also assumes you are familiar with the basic ideas and methods of hypothesis testing and confidence intervals. I am assuming you have been introduced to both these concepts in one of the prerequisite courses.