Study Guide for the General Applied Statistics Comprehensive Examination

Part II covers material from MAT 456 and from MAT 528 or MAT 526.

Review the homework assignments, activities, exams and quizzes you had.

I. MAT 456 Applied Regression Analysis

Student will be examined in the following broad topics:

- Simple linear regression
- Inferences about regression parameters
- derive simple linear regression parameter estimates
- Diagnostics for regression models
- Estimation for parameters
- Model selection and validation
- Analyze and interpret printouts for proposed models and test them for fit.
- Read the outputs from SAS
- When to use transformations

The afore-mentioned topics represent the first 11 chapters of the required textbook "Applied Linear Regression Models" by M. Kutner, C. Nachtsheim, and J. Neter, 4th edition.

The level of difficulty of the exam questions will be similar to that of the exams, homework assignments in the textbook and examples in lecture notes.

Specific references to the lecture notes are given below:

- Chapter 1: Linear regression with one predictor variable
  - Introduction to linear regression
- Chapter 2: Inferences in regression analysis
  - Inferences concerning β_1
  - Inferences concerning β_0
  - Some considerations on making inferences concerning β_0 and β_1
  - Interval estimation of $E(Y_h)$
  - Prediction of a new observation
  - Analysis of variance approach to regression analysis
  - General linear test approach
  - Descriptive measures of association between X and Y in the regression model
  - Considerations in applying regression analysis

- Chapter 3: Diagnostics and remedial measures
  - Diagnostics for predictor variable
  - Residuals

- Diagnostics for residuals and transformations
- Tests for constancy of error variance

- Chapter 4: Joint estimation of β_0 and β_1
  - Simultaneous estimation of mean responses
  - Simultaneous prediction intervals for a new observation
  - Inverse prediction

- Chapter 6: Multiple regression I
  - General linear regression model in matrix terms
  - Estimation of regression coefficients
  - Fitted values and residuals
  - Inferences about regression parameters
  - Analysis of variance results
  - Estimation of mean response and prediction of new observation
  - Diagnostics and remedial measures

- Chapter 7: Multiple regression II
  - Extra sums of squares
  - Uses of extra sums of squares in tests for regression coefficients
  - Coefficients of partial determination
  - Multicollinearity and its effects

- Chapter 8: Regression models for quantitative and qualitative predictors
  - Polynomial regression models
  - Interaction regression models
  - Qualitative predictors

- Chapter 9: Building the regression model I: model selection and validation
  - Criteria for model selection
  - Automatic search procedures for model selection
  - Model validation

- Chapter 10: Building the regression model II: diagnostics
  - Model adequacy for a predictor variable
  - Identifying outliers
  - Identifying influential cases

- Chapter 11: Building the regression model III: remedial measures and validation
  - Unequal error variance remedial measures
  - Multicollinearity remedial measures
  - Remedial measures for influential cases

Choose one of the two MAT 528 or MAT 526

II. MAT 528: Design and Analysis of Experiment

Review the following:

- Reasons for Randomization and use of blocking in designing scientific experiments.
- Latin Square Designs
- Factorial Designs
- Meaning and uses of confounding
- Role of aliases
- Computing EMS for a given model for Mixed and Random Models
- Analyses of Models from given ANOVA tables
- Fractional Factorial Designs and their uses

The afore-mentioned topics represent the first 11 chapters of the required textbook "Design and Analysis of experiments" by Douglas C. Montgomnery, 8th edition.

The level of difficulty of the exam questions will be similar to that of your exams, the homework assignments in the textbook or examples in lecture notes.

Specific references to the lecture notes are given below:

- Lecture 1: Introduction
  - Three principals of experimental design
  - Independent t test and paired t test

- Lecture 2: Compare n population means
  - Complete randomized block design
  - Statistical model and inference of CBD
  - Check model assumptions
  - Contrasts and orthogonal contrasts
  - Multiple testing and pairwise comparison

- Lecture 3: Block design, Latin square design
  - Randomized complete block design
  - Statistical model and inference of RCBD, ANOVA table
  - Check model assumptions, turkey's additive test for model additivity
  - RCBD with replication
  - Latin squares design and it's statistical model

- Lecture 4: Balanced incomplete block design and two factor factorial design
  - Balanced incomplete block design
  - Statistical model and inference of BIBD, ANOVA table
  - Check model assumptions
  - Statistical model of two factor factorial design and its inferences

- Lecture 5: 2^k factorial design
    - Introduction to 2^k factorial design
    - Statistical model and inferences use the contrast table for 2^2 factorial design

- Lecture6: Fractional factorial design
    - Blocking in 2^k factorial design
    - Confounding block effect in 2^k factorial design
    - Partial confounding
    - Fundamental principles for factorial effects

- Lecture7: Experiments with random effects
    - CRD with random effects
    - Statistical model of CRD with random effects and inferences (Confidence intervals, ICC)
    - Two factor mixed effect model and its inferences

III. MAT 526 Sampling Portion

Students will be examined in the following broad topics:

- Basic concepts of sampling survey
- Concepts of different sampling methods
- Statistical analysis of sampling survey data using different sampling method

The afore-mentioned topics represent the six lecture notes covered in the course MAT 526. Discussion of these topics can be found in the book "Elementary Survey Sampling (7^{th} edition)" by Richard L. Scheaffer, et al.

The questions will be drawn for the comprehensive exam. Questions may have multiple parts. The level of difficulty will be similar to that of the weekly assignment problems, in class examples, or problems in tests and final exam.

Students are expected to spend no more than 75 working on the MAT 526 component of the comprehensive exam.

Specific references to the lecture notes are given below:

- Lecture Note 1: Introduction and Basic Concepts
    - Basic concepts of sampling survey
- Lecture Note 2: Simple Random Sampling
    - Basic concepts of simple random sampling (SRS)
    - Estimation of population mean with bound/margin of error
    - Estimation of population total with bound/margin of error

- Estimation of population proportion with bound/margin of error
- Determine the sample size with specified estimation purpose
- Compare the means of two populations
- Compare the proportions between two independent polls
- Compare the proportions within a single poll

- Lecture Note 3: Systematic Sampling
  - Basic concepts of systematic sampling
  - Systematic sampling compared to simple random sampling
  - Estimation of population mean with bound/margin of error
  - Estimation of population total with bound/margin of error
  - Estimation of population proportion with bound/margin of error
  - Determine the sample size with specified estimation purpose
- Lecture Note 4*:* Stratified Random Sampling
  - Basic concepts and advantages of stratified random sampling
  - Estimation of population mean with bound/margin of error
  - Estimation of population total with bound/margin of error
  - Estimation of population proportion with bound/margin of error
  - Determine the sample size with specified estimation purpose
  - Optimal allocation under fixed cost or fixed error tolerance
  - Post stratification
- Lecture Note 5*:* Ratio Estimation
  - Basic concepts and advantages of ratio estimation
  - Estimations under simple random sampling
    - Estimation of ratio with bound/margin of error
    - Estimation of population mean with bound/margin of error
    - Estimation of population total with bound/margin of error
    - Estimation of population proportion with bound/margin of error
  - Estimations under stratified sampling
    - Estimation of separate/combined ratio with bound/margin of error
    - Estimation of population mean with bound/margin of error
    - Estimation of population total with bound/margin of error
    - Estimation of population proportion with bound/margin of error
- Lecture Note 6*:* Regression Estimation and Difference Estimation
  - Basic concept and advantages of regression/difference estimation
  - Estimation of population mean with bound/margin of error using regression estimation/difference estimation
  - Relative efficiency of two estimators