

MS in Applied Statistics: Study Guide for the Biostatistics Concentration Comprehensive Examination

The Part II comprehensive examination is a three-hour closed-book exam that is offered on the second Saturday of the autumn and spring quarters.

This exam covers materials in MAT 456, and 421 or 425.

1. MAT 456 Applied Regression Analysis

Student will be examined in the following broad topics:

- Simple linear regression
- Inferences about regression parameters
- 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:

1. Chapter 1: Linear regression with one predictor variable

- Introduction to linear regression

2. 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

3. **Chapter 3: Diagnostics and remedial measures**
 - Diagnostics for predictor variable
 - Residuals
 - Diagnostics for residuals and transformations
 - Tests for constancy of error variance

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

5. **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

6. **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

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

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

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

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

2. MAT 421 Basic Biostatistics

Student will be examined in the following broad topics:

- Types of data
- Probability
- Population vs. sample
- Discrete and continuous distributions
- Inferences about population parameters
- Association test
- Regression

The afore-mentioned topics represent the 9 lecture notes covered in the course MAT 421.

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

Specific references to the lecture notes are given below:

1. Lecture 1: Introduction to Biostatistics

- Variables
- Variety of types of data
- Graphical and numerical ways of summarizing data

2. Lecture 2: Introduction to Probability

- a. Ideas of chance
- b. Basic of calculating probabilities

3. Lecture 4: Bernoulli and Binomial Distributions

- a. Pattern of discrete outcomes
- b. Result of a multiple number of Bernoulli trials

4. Lecture 5: Normal Distribution

- a. Pattern of outcomes that have values on a continuum
- b. Normal and standard normal distributions
- c. Normal probability distribution probabilities

5. Lecture 6: Estimation

- a. Point and confidence interval estimates from single data
- b. Point and confidence interval estimates from paired data
- c. Point and confidence interval estimates from two independent data

6. Lecture 7: Hypothesis Testing

- a. Research question to testable hypotheses
- b. Steps in performing a statistical hypothesis test
- c. Interpretation of p-value with respect to rejection of non-rejection

7. Lecture 8: Chi Square Tests

- a. Association for a 2×2 table
- b. Association for a $r \times c$ table
- c. Interpretation of association

3. MAT 425 Survival Analysis

You should be able to know

- how to compute percentile using KM and graphs
- the difference among different types of censoring
- how to estimate the hazard and survival functions from given distributions
- how to calculate relative risks of individuals
- how to construct confidence interval for the risk of death for patients in stage IV relative to stage I
- the difference between hazard function and survival function
- compute logrank for a given small data sets
- the difference among the different weights for the logrank tests
- the different techniques for estimating hazard function (bandwidth, kernel functions)

Textbook: Klein and Moeschberger (KM) Survival Analysis: Techniques for Censored and Truncated Data, Springer, New York, 2003.

The level of difficulty of the exam questions will be similar to that of exam1, final exam, the homework assignments and examples in lecture notes.

1. Chapter 4 Basic Quantities and models

- The survival and hazard functions
- The mean residual life function and median life
- Common parametric models for survival data
- Regression Models for Survival Data

2. Chapter 3 Censoring and Truncation

- Right censoring

- Left and interval censoring
- Truncation
- Likelihood Construction for Right Censored Data

3. Chapter 4 Nonparametric estimation of basic quantities for right censored data

- Estimation of survival and cumulative hazard functions for right censored data
- Point estimates for the mean and median survival time
- Pointwise Confidence Intervals for the Survival Function

4. Chapter 6 Topic in univariate estimation

- Estimating the hazard function using Kernel and KM type methods

5. Chapter 7 Hypothesis Testing

- One sample test
- Tests for two or more samples (Logrank test)

6. Chapter 8 Semiparametric Proportional Hazards Regression

- Cox proportional regression
- Coding covariates
- Model building using the proportional hazards model (interpret the coefficients and construct confidence intervals for the estimates)
- Know how to interpret the risk coefficients of the proportional model
- Estimation of the Survival Function