Statistics for Risk Modeling Exam—January 2027

The Statistics for Risk Modeling exam is a three and one-half hour exam that consists of 35 multiple-choice questions and is administered as a computer-based test (CBT). For additional details, please refer to Exam Rules.

The goal of the syllabus for this examination is to provide an understanding of the basics of several important analytic methods. This exam is a prerequisite for the Predictive Analytics exam, which will go deeper into each of the covered techniques.

A thorough knowledge of calculus, probability (as covered in the P Exam) and mathematical statistics (as covered in the VEE Mathematical Statistics) is assumed.

The following learning objectives are presented with the understanding that candidates are allowed to use specified calculators on the exam.

A variety of tables is available below for the candidate and will be provided to the candidate at the examination. These include values for the standard normal distribution, *t* distribution and chi-square distribution. This information will be available as a pdf for the CBT exam candidates found under an **Exhibit** button during the exam. Paper/pencil candidates will receive a hard copy of the table(s). Since the table(s) will be included with the examination no matter which format is taken, candidates will not be allowed to bring copies of the tables into the examination room.

Please check the Updates section on this exam's home page for any changes to the exam or syllabus.

In the learning outcomes, weights have been provided to indicate the relative emphasis on different sections. The ranges of weights shown are intended to apply to the large majority of exams administered. On occasion, the weights of topics on an individual exam may fall outside the published range. Candidates should also recognize that some questions may cover multiple learning outcomes.

For this exam, ability to solve problems using the R programming language will not be assumed. However, questions may present R output for interpretation.

Each multiple-choice problem includes five answer choices identified by the letters A, B, C, D, and E, only one of which is correct. Answers for some questions have been rounded.

As part of the computer-based testing process, a few pilot questions will be randomly placed in the exam (both paper and pencil and computer-based forms). These pilot questions are included to judge their effectiveness for future exams, but they will NOT be used in the scoring of this exam. All other questions will be considered in the scoring. All unanswered questions are scored incorrect. Therefore, candidates should answer every question on the exam. There is no set requirement for the distribution of correct answers for the multiple-choice preliminary examinations. It is possible that a particular answer choice could appear many times on an examination or not at all. Candidates are advised to answer each question to the best of their ability, independently from how they have answered other questions on the exam.

Since the CBT exam will be offered over a period of a few days, each candidate will receive a test form composed of questions selected from a pool of questions. Statistical scaling methods are used to ensure within reasonable and practical limits that, during the same testing period of a few days, all forms of the test

are comparable in content and passing criteria. The methodology that has been adopted is used by many credentialing programs that give multiple forms of an exam.

Unofficial pass/fail results will be sent within one hour to the email address you used to schedule your appointment.

LEARNING OBJECTIVES

1. **Topic: Basics of Statistical Learning (5-10%)**
Learning Objectives
The Candidate will understand key concepts of statistical learning.
Learning Outcomes
The Candidate will be able to: a) Define terms used to classify the types of modeling problems and methods, including supervised versus unsupervised learning and regression versus classification. b) Compare the common methods of assessing model accuracy. c) Understand how the bias-variance tradeoff impacts the selection of statistical learning methods. d) Understand resampling methods used for model validation, including • Training set vs. test set approach • *k*-fold cross-validation • Leave-one-out cross-validation

2. Topic: Linear Models (40-50%)

Learning Objectives

The Candidate will understand key concepts concerning generalized linear models.

Learning Outcomes

The Candidate will be able to:

a) Compare model assumptions for ordinary least squares and generalized linear models.

b) Identify the members of the exponential family of distributions and corresponding link functions.

c) Apply the business context of a problem to interpret parameters.

d) Interpret diagnostic tests of model fit and assumption checking, using

- Graphical methods
- Quantitative methods

e) Select an appropriate model, considering

- Distributions and link functions
- Variable transformations and interactions
- t and F tests
- AIC and BIC
- Likelihood ratio test

f) Calculate and interpret predicted values, and confidence and prediction intervals.

g) Understand how approaches may differ compared to using an ordinary least squares model, including

- Regularized regression (lasso, ridge regression)
- *K*-nearest neighbors

3. Topic: Time Series Models (10-15%)

Learning Objectives

The Candidate will understand key concepts concerning regression-based time series models.

Learning Outcomes

The Candidate will be able to:

a) Define the concepts and components of stochastic time series processes, including random walks, stationarity, and autocorrelation.

b) Describe specific time series models, including, exponential smoothing, autoregressive, and autoregressive conditionally heteroskedastic models.

c) Calculate and interpret predicted values and confidence intervals.

4. Topic: Decision Trees (20-25%)

Learning Objectives

The Candidate will understand key concepts concerning decision tree models.

Learning Outcomes

The Candidate will be able to:

a) Describe the construction of decision trees, including
   - How they are optimally fit to training data
   - How they are pruned to mitigate overfitting

b) Predict outcomes using
   - Classification trees
   - Regression trees

c) Describe bagging, boosting, and random forests and the hyperparameters used to control them.

d) Compare decision trees to linear models including uses and relative strengths.

5. Topic: Unsupervised Learning Techniques (10-15%)
Learning Objectives
The Candidate will understand key concepts concerning principal component and cluster analysis.
Learning Outcomes
The Candidate will be able to: a) Define principal components, including how they are calculated. b) Interpret the results of a principal components analysis, considering loading factors and proportion of variance explained. c) Describe and compare the algorithms for: • *K*-means clustering • Hierarchical clustering d) Explain methods for deciding the number of clusters.

Textbooks

Regression Modeling with Actuarial and Financial Applications, Edward W. Frees, 2010, New York: Cambridge. ISBN: 978-0521135962.

Chapter 1 – Background only
Chapter 2 – Sections 1-8
Chapter 3 – Sections 1-5
Chapter 5 – Sections 1-7
Chapter 6 – Sections 1-3
Chapter 7 – Sections 1-6
Chapter 8 – Sections 1-4
Chapter 9 – Sections 1-5
Chapter 11 – Sections 1-6
Chapter 12 – Sections 1-4
Chapter 13 – Sections 1-6

An Introduction to Statistical Learning, with Applications in R Second Edition, James, Witten, Hastie, Tibshirani, New York: Springer. A PDF of the text can be obtained via the link found here: https://www.statlearning.com. Scroll to the bottom of the page and click on the box labeled "2nd Edition of ISLR with R."

Chapter 2 – Sections 1-3
Chapter 3 – Sections 1-6
Chapter 5 – Sections 1 and 3 (excluding 5.3.4)
Chapter 6 – Sections 1-5
Chapter 8 – Sections 1-3 (excluding 8.2.4, 8.2.5, and 8.3.5)
Chapter 12 – Sections 1-2, 4-5 (excluding 12.5.2)

While exercises are not included in the required readings, candidates are encouraged to work them as part of the learning experience.

OTHER RESOURCES:

Tables for Exam SRM

Sample Questions and Solutions

A NOTE ON AIC AND BIC:

Two commonly used measures of model goodness-of-fit are Akaike information criterion (AIC) and Bayesian information criterion (BIC). While these measures are common, there are multiple formulas for both measures. Some of these alternate formulas eliminate a constant term, and some are special cases that only apply in certain contexts.

For Exam SRM, we will only use AIC and BIC in the context of ordinary linear models, and we will use the formulas found in Section 6.1.3 of *Introduction to Statistical Learning*. For clarity, these formulas are repeated below.

$$\text{AIC} = \frac{1}{n}(\text{RSS} + 2d\hat{\sigma}^2) \qquad \text{BIC} = \frac{1}{n}(\text{RSS} + \ln(n)d\hat{\sigma}^2)$$