Learning Objectives Exam 2 - Data Management and Exploratory Data Analysis and Visualization

Data Concepts

A. Accessing Raw Data

MODULE	LEARNING OBJECTIVES
Internal Sources of Data - Operational data	Understand what typical attributes are made available in each of the following data sources: • Policy and Premium Data • Claims Information • Claim Notes • Billing Information • Producer Information Understand how corrections for each of these attributes are recorded. Understand timing of collection and updating. Understand how quality can change over time.
Internal Sources of Data - Reported data	Understand why insurance companies produce statistical files, typical attributes in stat files and the advantages and disadvantages of using stat files over operational data.
External Sources of data - Non Insurance	Understand how to access and the uses of external sources of data including: • Demographic information • Customer Financial Information • Business Financial Information • Behavioral data • Driving Records and Motor Vehicle Reports • Government sources Understand who collects the information, for what purposes, how frequently it is updated and how it is distributed. For each of these sources understand what typical attributes are made available. Understand various derived attributes. Understand if there is a clear way to merge the data into databases used for analysis.
External Sources of data - Insurance Specific	Understand External sources of Insurance Specific data such as: • Historical claims reports • Industry trend factors • Loss Development factors • Data available in NCCI Statistical Bulletin

External Sources of data - Other	Be aware of other external sources of data such as: • Social media feeds • Web APIs • ACID • NoSQL databases

B. Accessing Raw Data

MODULE	LEARNING OBJECTIVES
Broad classification of data	Understand the broad classifications of data: quantitative vs qualitative; nominal, ordinal, interval continuous; transactional, snapshots, aggregated. Understand transaction and snapshot data and how to combine snapshots from various times into transaction files.
Data Types	Understand the various data types: Numeric, string, date, geographic. Understand how to read, store and display each data type. Understand issues with Date types, such as different formatting across different data sources. Understand Unicode coding and why it is used. Understand structured versus unstructured data and the various forms of unstructured data such as document, map, voice, and image.
Data File Formats	Be able to read and write each of the following file formats into one or more data tables: Formatted text, Delimited text, Relational databases, Excel files, R and Python data frames. Know how to parse string, numeric and date fields. Understand how to read Unicode data. Understand data warehouses and data marts for storing data. Understand XML, HTML and JSON formats and web scraping.
Data Profiling	Demonstrate ability to profile data including: inspecting rows of data, read data catalogs and metadata. Create descriptive statistics and graphs that profile the data.
Dealing with messy data	Be able to detect and remediate missing, miscoded and anomalous data. Understand sampling bias and clustering of values.

Concept of Test Data	Be able to work with small test data and create sample data using simple filters. Be able to sample from related tables.
Basics of querying data	Be able to query data, including use of SQL for queries and data management. Read first few rows or columns. Be able to filter rows, create derived columns. Be able to summarize data by groups and produce aggregated data. Know how to do the followings Joins: Union (concatenation), Left Join, Right Join, Inner Join, Outer Join, join multiple tables. Understand the SQL grammar. Be able to mitigate SQL performance issues. Understand SQL indexes. Be able to write SQL functions.

C. Working with Data

MODULE	LEARNING OBJECTIVES
Basics of string processing	Understand regular expressions, Fuzzy Matching and Hashing. Know how to use regular expressions to search for patterns in strings. Use look ahead and look behind concepts. Be able to search for and replace strings including use of back reference. Be able to do approximate string matching. Understand hashing, various hashing algorithms and uses of hashing.
Simple Insurance data preparation applications	Know how to prepare data for an insurance project. Know how to use of premium, loss, policy information and external data sources. Be able to aggregate transaction level, claim level, etc. data and join the data sources. Do simple profiling of the combined data.

D. Data Quality

MODULE	LEARNING OBJECTIVES
Data Quality	Summarize concepts of data quality. The impact of data on actuarial work and projects
Principles of Data Quality	Given a principle of data quality, provide an example that illustrates the principle

Data Quality ASOP 23	Given a concept from ASOP 23, Provide examples of application and use
Life Cycle for insurance data	For each step in the life cycle for insurance data describe the purpose, responsible parties and errors typically encountered
Metadata	Summarize metadata including: • How metadata are defined, • The actuary/analyst's role in creating and sharing metadata, • How metadata are shared across organizations and • The data collected under different statistical plans
The need for Aggregate Insurance statistical data	Explain the regulator and business needs for statistical data

E. Insurance Applications

MODULE	LEARNING OBJECTIVES
Creating modeling datasets for underwriting model	Be able to create a modeling dataset for an underwriting model that will use policy, claims, beginning of term financial information, demographic and external data. Determine what each row in the modeling data will represent (example: Policy term, Coverage Exposure term) Understand potential target and predictor variables. Determine if adjustments such as development, trend or on-leveling of premium, are needed.
Creating modeling datasets for claims model	Be able to create a model dataset for claims applications. Determine target and predictor variables. Adjust data for development, trend, etc., as needed. Understand what each row in the modeling data will represent (example: Claim data at each evaluation month). Consider the following types of variables: policy characteristics, Policy's historic claim characteristics, demographic variables, financial information for policyholder and claimant. Be familiar with the following applications: First Notice of Loss Model, Fraud detection model, Claims Complexity model and other applications.

F. Regulation

MODULE	LEARNING OBJECTIVES
Regulation of data	Understand Regulation of data such as privacy regulation (HIPPA) Fair Credit Reporting, etc.

Data Tools, Exploration and Visualization

A. Univarate Exploratory Data Analysis

MODULE	LEARNING OBJECTIVES
Overall Objective	Understand and be able to apply the univariate descriptive statistics and graphs and be familiar with how they can be used to characterize data. Be able to use univariate statistics to detect outliers and data errors. Understand how to use graphs for transformation of data.
Perform Exploratory Data Analysis (EDA) with Descriptive Statistics	Find central estimate, measures of dispersions and measures of shape. Use 5 (or more) point summary to characterize data and detect errors. Understand difference between an outlier and an error. Understand how to address extreme values with truncation, censoring, etc.
Perform graphical analysis on numeric data	Be able to apply Histograms, boxplots, Kernel Density estimates and QQ plots. Determine bins for histograms.
Perform analysis on categorical data.	Be able to use Bar plots and Categorical tables to display categorical data and detect data anomalies and sparse data. Be able to bin categories.
Perform analysis of time series data	Use methods for graphically displaying time series data. Use graphs and statistics to assess time series correlations.

Understand smoothing	Distinguishing pattern from noise. Understanding why smoothing/whitening and normalization are performed and how they are applied

B. Multivariate Exploratory Data Analysis

MODULE	LEARNING OBJECTIVES
Overall Objective	Understand how to use multivariate summaries and displays to uncover relationships in data, to detect outliers and to formulate preliminary hypotheses
Multivariate descriptive statistics	Know how to assess linear and non-linear correlations, crosstabulations and pivot table multi-way summaries of data
Multivariate graphics	Be able to use common multivariate graphical EDA tools including: scatterplots, scatterplot matrices, multi-way panel/lattice plots for boxplots, histograms, etc.

C. Visualization

MODULE	LEARNING OBJECTIVES
Theory of Visualization	Understand the overall philosophy of data visualization and information display including: less is more, moral responsibility in use of charts, dollars vs percent change, and choice of base.
Design of graphs	Understand the use of color, RGB vs. HSV, selecting palettes, dealing with colorblindness, perfectly perceptually-uniform color maps. Apply marker style and line style, gridlines and background images and shading. Perform axis design and labeling. Be able to combine multiple plots on one axis or groups of axes. Design text labels, including axes labels and legends. Enhance graphs with text. Understand proper use of original vs log scale. Understand use of alpanumerics in graphs.
Tabular visualization	Be able to design informative tables. Understand how to drill down on "live" versus static data.

Dashboards	Understand how to create informative dashboards including combining multiple graphical elements, design of user interactions and choice elements.
Diagrams	Understand the principles of diagramming including the following types of diagrams: Flow Charts, Using plotting data as an image, Word clouds and Network graphs.
Data preparation for visualization	Understand data preparation and transformations needed to supply data to graphics packages, e.g. by-variables as columns or data elements in single column. Understand data consolidation for graphing.
Audience and Purpose	Understand how to consider the audience and the purpose when designing presentations including technical vs non-technical audience, familiarity with the subject matter; whether it's purpose is sales, information, education, entertainment, etc. Consider whether presentation must be understood as a stand-alone or will accompany an explanation.