

## SAS® Interactive Model Building using SAS® Visual Statistics 8.4 on SAS Viya

### SAS® Visual Statistics Cross-functional Tasks

#### Prepare data using SAS® Visual Analytics.

- Manage explorations and visualizations.
- Impute a variable.
- Transform a variable.
- Create an aggregated measure.
- Replace dirty data with missing values.
- Combine multiple categories into fewer levels.

#### Filter data used for a model.

- Exclude selections to filter data.
- Apply filters to visualization and data source.
- Review Measure Details.

#### Use interactive group-by.

- Explain group-by modeling.
- Assign a group-by variable to a predictive model (logistic regression, linear regression model and generalized linear model).
- Interactively examine the Fit Summary for group-by models.
- Choose the best fitting group-by model using fit statistics and Variable Importance.
- Interpret model results using advanced group-by feature.
- Examine the summary table for group-by processing.

#### Perform model validation

- Explain model complexity.
- Create and assign a partition.
- Explain model selection based off partitioned data.
- Choose the best fitting model with partitioned data.
- Interpret model results with partitioned data.

## Building and Assessing Segmentation Models

### Perform unsupervised segmentation using cluster analysis.

- Explain unsupervised classification.
- Given a scenario, set proper inputs for k-means algorithm.
- Build a cluster analysis in SAS® Visual Statistics.
- Assign roles for cluster analysis.
- View and edit cluster properties.
- Set Parallel Coordinate properties for a cluster.
- Given a scenario, appropriately change the number of clusters.
- Derive a cluster ID variable and use it in another visualization.

### Analyze cluster results.

- Interpret a Cluster Matrix.
- Interpret Parallel Coordinates plot.
- Interpret Cluster Summary tab.

### Perform supervised segmentation using decision trees.

- Explain how split points are determined.
- Assign variable roles for a decision tree.
- Define decision tree properties.
- Describe how predictions are formulated for a decision tree.
- Explain variable selection methods for decision trees.
- Derive a leaf ID for use in other models.
- Prune a decision tree.

### Asses decision tree results.

- Interpret tree with Tree Map.
- Interpret Leaf statistics.
- Interpret Assessment panel.
- Investigate leaf nodes.
- Explain icicle plot.

## Building and Assessing Regression-type Models

### Explain linear models.

- Explain linear regression.
- Model effects usage.
- Given a scenario, determine when to use a linear regression model vs. a generalized linear model.

### Perform linear regression modeling.

- Assign linear regression roles.
- Add Interaction Effect.
- Define linear regression properties.
- Explain informative missingness.
- Review outlier details and exclude outliers.

### Perform generalized linear regression modeling.

- Assign generalized linear model roles.
- Assign offset variable.
- Define linear regression properties.
- Link functions and distributions in generalized linear models.
- Given a scenario, choose appropriate distribution and link function.

### Perform logistic regression modeling.

- Explain logistic regression essentials.
- Explain prediction in logistic regression.
- Explain variable selection in SAS® Visual Statistics.
- Specify which variable is the event (binary).
- Specify how a multinomial response variable is used in SAS® Visual Statistics.
- Assign logistic regression roles.
- Define logistic regression properties.
- Specify when to use appropriate link function when building a predictive model.

### Assess model results.

- Interpret Fit Summary window.
- Interpret Residual Plot.
- Interpret ROC chart (KS Statistic).
- Evaluate Misclassification plot.
- Evaluate the Lift chart.

- Interpret Influence plot.
- Interpret Summary bar.
- Assess residuals and other model diagnostics to choose an appropriate distribution and link function.
- Derive predicted values and describe in terms of predicted probabilities in SAS® Visual Statistics.
- Apply prediction cut-off.

### **Perform generalized additive modeling**

- Explain generalized additive model essentials.
- Explain spline term essentials.
- Explain differences and similarities between generalized additive models and generalized linear regression.
- Explain advantages and disadvantages of generalized additive models.
- Assess model fit statistics that are common to generalized additive models (GCV, UBRE).

### **Perform nonparametric logistic regression modeling**

- Explain nonparametric logistic regression essentials.
- Explain differences between nonparametric logistic regression and logistic regression.
- Explain advantages and disadvantages of nonparametric logistic regression.

## **Model Comparison and Scoring**

### **Compare models**

- Explain model comparison features.
- Assign model comparison properties.
- Interpret comparison results using Assessment panel, Fit Statistics, ROC charts, concordance statistics, misclassification, etc.
- Interpret Summary Table for model comparison (statistics, variable importance).
- Given a scenario, use a particular fit statistic to select a champion model.
- Define the conditions that make models comparable in SAS® Visual Statistics.

### **Score models**

- Explain scoring functionality.
- Export score code.
- Implement score code.
- Identify which SAS® tools can score new data using score code generated by SAS® Visual Statistics.

**Note:** All 14 main objectives will be tested on every exam. The 80 expanded objectives are provided for additional explanation and define the entire domain that could be tested.