SAS A00-240

SAS STATISTICAL BUSINESS ANALYST CERTIFICATION QUESTIONS & ANSWERS

Exam Summary – Syllabus – Questions

A00-240

SAS Certified Statistical Business Analyst Using SAS 9 - Regression and Modeling
60 Questions Exam – 68% Cut Score – Duration of 120 minutes

www.AnalyticsExam.Com

Table of Contents

Know Your A00-240 Certification Well:

The A00-240 is best suitable for candidates who want to gain knowledge in the SAS Advanced Analytics. Before you start your A00-240 preparation you may struggle to get all the crucial SAS Statistical Business Analyst materials like A00-240 syllabus, sample questions, study guide.

But don't worry the A00-240 PDF is here to help you prepare in a stress free manner.

The PDF is a combination of all your queries like-

- What is in the A00-240 syllabus?
- How many questions are there in the A00-240 exam?
- Which Practice test would help me to pass the A00-240 exam at the first attempt?

Passing the A00-240 exam makes you SAS Certified Statistical Business Analyst Using SAS 9 - Regression and Modeling. Having the SAS Statistical Business Analyst certification opens multiple opportunities for you. You can grab a new job, get a higher salary or simply get recognition within your current organization.

A00-240 SAS Statistical Business Analyst Certification Details:

Exam Name	SAS Certified Statistical Business Analyst Using SAS 9
Exam Code	A00-240
Exam Duration	120 minutes
Exam Questions	60
Passing Score	68%
Exam Price	$180 (USD)
Training	**Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression** **Predictive Modeling Using Logistic Regression**
Book	**SAS® Certification Prep Guide: Statistical Business Analysis Using SAS®9**
Exam Registration	**Pearson VUE**
Sample Questions	**SAS Statistical Business Analyst Certification Sample Question**
Practice Exam	**SAS Statistical Business Analyst Certification Practice Exam**

A00-240 Syllabus:

Objective	Details
ANOVA - 10%	
Verify the assumptions of ANOVA	- Explain the central limit theorem and when it must be applied - Examine the distribution of continuous variables (histogram, box -whisker, Q-Q plots) - Describe the effect of skewness on the normal distribution - Define H0, H1, Type I/II error, statistical power, p-value - Describe the effect of sample size on p-value and power - Interpret the results of hypothesis testing - Interpret histograms and normal probability charts - Draw conclusions about your data from histogram, box-whisker, and Q-Q plots - Identify the kinds of problems may be present in the data: (biased sample, outliers, extreme values) - For a given experiment, verify that the observations are independent - For a given experiment, verify the errors are normally distributed - Use the UNIVARIATE procedure to examine residuals - For a given experiment, verify all groups have equal response variance - Use the HOVTEST option of MEANS statement in PROC GLM to asses response variance
Analyze differences between population means using the GLM and TTEST procedures	- Use the GLM Procedure to perform ANOVA • CLASS statement • MODEL statement • MEANS statement • OUTPUT statement - Evaluate the null hypothesis using the output of the GLM procedure - Interpret the statistical output of the GLM procedure (variance derived from MSE, F value, p-value $R^{**}2$, Levene's test) - Interpret the graphical output of the GLM procedure - Use the TTEST Procedure to compare means
Perform ANOVA post hoc test to evaluate treatment effect	- Use the LSMEANS statement in the GLM or PLM procedure to perform pairwise comparisons - Use PDIFF option of LSMEANS statement - Use ADJUST option of the LSMEANS statement (TUKEY and DUNNETT) - Interpret diffograms to evaluate pairwise comparisons - Interpret control plots to evaluate pairwise comparisons

Objective	Details
	- Compare/Contrast use of pairwise T-Tests, Tukey and Dunnett comparison methods
Detect and analyze interactions between factors	- Use the GLM procedure to produce reports that will help determine the significance of the interaction between factors. MODEL statement - LSMEANS with SLICE=option (Also using PROC PLM) - ODS SELECT - Interpret the output of the GLM procedure to identify interaction between factors: - p-value - F Value - R Squared - TYPE I SS - TYPE III SS
	Linear Regression - 20%
Fit a multiple linear regression model using the REG and GLM procedures	- Use the REG procedure to fit a multiple linear regression model - Use the GLM procedure to fit a multiple linear regression model
Analyze the output of the REG, PLM, and GLM procedures for multiple linear regression models	- Interpret REG or GLM procedure output for a multiple linear regression model: convert models to algebraic expressions - Convert models to algebraic expressions - Identify missing degrees of freedom - Identify variance due to model/error, and total variance - Calculate a missing F value - Identify variable with largest impact to model - For output from two models, identify which model is better - Identify how much of the variation in the dependent variable is explained by the model - Conclusions that can be drawn from REG, GLM, or PLM output: (about H0, model quality, graphics)
Use the REG or GLMSELECT procedure to perform model selection	- Use the SELECTION option of the model statement in the GLMSELECT procedure - Compare the differentmodel selection methods (STEPWISE, FORWARD, BACKWARD) - Enable ODS graphics to display graphs from the REG or GLMSELECT procedure - Identify best models by examining the graphical output (fit criterion from the REG or GLMSELECT procedure) - Assign names to models in the REG procedure (multiple model statements)
Assess the validity of a given regression	- Explain the assumptions for linear regression - From a set of residuals plots, asses which assumption

Objective	Details
model through the use of diagnostic and residual analysis	about the error terms has been violated - Use REG procedure MODEL statement options to identify influential observations (Student Residuals, Cook's D, DFFITS, DFBETAS) - Explain options for handling influential observations - Identify collinearity problems by examining REG procedure output - Use MODEL statement options to diagnose collinearity problems (VIF, COLLIN, COLLINOINT)
Logistic Regression - 25%	
Perform logistic regression with the LOGISTIC procedure	- Identify experiments that require analysis via logistic regression - Identify logistic regression assumptions - logistic regression concepts (log odds, logit transformation, sigmoidal relationship between p and X) - Use the LOGISTIC procedure to fit a binary logistic regression model (MODEL and CLASS statements)
Optimize model performance through input selection	- Use the LOGISTIC procedure to fit a multiple logistic regression model - LOGISTIC procedure SELECTION=SCORE option - Perform Model Selection (STEPWISE, FORWARD, BACKWARD) within the LOGISTIC procedure
Interpret the output of the LOGISTIC procedure	- Interpret the output from the LOGISTIC procedure for binary logistic regression models: Model Convergence section - Testing Global Null Hypothesis table - Type 3 Analysis of Effects table - Analysis of Maximum Likelihood Estimates table - Association of Predicted Probabilities and Observed Responses
Score new data sets using the LOGISTIC and PLM procedures	- Use the SCORE statement in the PLM procedure to score new cases - Use the CODE statement in PROC LOGISTIC to score new data - Describe when you would use the SCORE statement vs the CODE statement in PROC LOGISTIC - Use the INMODEL/OUTMODEL options in PROC LOGISTIC - Explain how to score new data when you have developed a model from a biased sample
Prepare Inputs for Predictive Model Performance - 20%	
Identify the potential challenges when	- Identify problems that missing values can cause in creating predictive models and scoring new data sets - Identify limitations of Complete Case Analysis

Objective	Details
preparing input data for a model	- Explain problems caused by categorical variables with numerous levels - Discuss the problem of redundant variables - Discuss the problem of irrelevant and redundant variables - Discuss the non-linearities and the problems they create in predictive models - Discuss outliers and the problems they create in predictive models - Describe quasi-complete separation - Discuss the effect of interactions - Determine when it is necessary to oversample data
Use the DATA step to manipulate data with loops, arrays, conditional statements and functions	- Use ARRAYs to create missing indicators - Use ARRAYS, LOOP, IF, and explicit OUTPUT statements
Improve the predictive power of categorical inputs	- Reduce the number of levels of a categorical variable - Explain thresholding - Explain Greenacre's method - Cluster the levels of a categorical variable via Greenacre's method using the CLUSTER procedure • METHOD=WARD option • FREQ, VAR, ID statement • Use of ODS output to create an output data set - Convert categorical variables to continuous using smooth weight of evidence
Screen variables for irrelevance and non-linear association using the CORR procedure	- Explain how Hoeffding's D and Spearman statistics can be used to find irrelevant variables and non-linear associations - Produce Spearman and Hoeffding's D statistic using the CORR procedure (VAR, WITH statement) - Interpret a scatter plot of Hoeffding's D and Spearman statistic to identify irrelevant variables and non-linear associations
Screen variables for non-linearity using empirical logit plots	- Use the RANK procedure to bin continuous input variables (GROUPS=, OUT= option; VAR, RANK statements) - Interpret RANK procedure output - Use the MEANS procedure to calculate the sum and means for the target cases and total events (NWAY option; CLASS, VAR, OUTPUT statements) - Create empirical logit plots with the SGPLOT procedure - Interpret empirical logit plots

Objective	Details
Measure Model Performance - 25%	
Apply the principles of honest assessment to model performance measurement	- Explain techniques to honestly assess classifier performance - Explain overfitting - Explain differences between validation and test data - Identify the impact of performing data preparation before data is split
Assess classifier performance using the confusion matrix	- Explain the confusion matrix - Define: Accuracy, Error Rate, Sensitivity, Specificity, PV+, PV- - Explain the effect of oversampling on the confusion matrix - Adjust the confusion matrix for oversampling
Model selection and validation using training and validation data	- Divide data into training and validation data sets using the SURVEYSELECT procedure - Discuss the subset selection methods available in PROC LOGISTIC - Discuss methods to determine interactions (forward selection, with bar and @ notation) - Create interaction plot with the results from PROC LOGISTIC - Select the model with fit statistics (BIC, AIC, KS, Brier score)
Create and interpret graphs (ROC, lift, and gains charts) for model comparison and selection	- Explain and interpret charts (ROC, Lift, Gains) - Create a ROC curve (OUTROC option of the SCORE statement in the LOGISTIC procedure) - Use the ROC and ROCCONTRAST statements to create an overlay plot of ROC curves for two or more models - Explain the concept of depth as it relates to the gains chart
Establish effective decision cut-off values for scoring	- Illustrate a decision rule that maximizes the expected profit - Explain the profit matrix and how to use it to estimate the profit per scored customer - Calculate decision cutoffs using Bayes rule, given a profit matrix - Determine optimum cutoff values from profit plots - Given a profit matrix, and model results, determine the model with the highest average profit

SAS A00-240 Sample Questions:

Question: 1

An analyst has determined that there exists a significant effect due to region. The analyst needs to make pairwise comparisons of all eight regions and wants to control the experimentwise error rate.

Which GLM procedure statement would provide the correct output?

a) lsmeans Region / pdiff=all adjust=dunnett;
b) lsmeans Region / pdiff=all adjust=tukey;
c) lsmeans Region / pdiff=all adjust=lsd;
d) lsmeans Region / pdiff=all adjust=none;

Answer: b

Question: 2

An analyst builds a logistic regression model which is 75% accurate at predicting the event of interest on the training data set. The analyst presents this accuracy rate to upper management as a measure of model assessment.

What is the problem with presenting this measure of accuracy for model assessment?

a) This accuracy rate is redundant with the misclassification rate.
b) It is pessimistically biased since it is calculated from the data set used to train the model.
c) This accuracy rate is redundant with the average squared error.
d) It is optimistically biased since it is calculated from the data used to train the model.

Answer: d

Question: 3

The LOGISTIC procedure will be used to perform a regression analysis on a data set with a total of 10,000 records. A single input variable contains 30% missing records.

How many total records will be used by PROC LOGISTIC for the regression analysis?
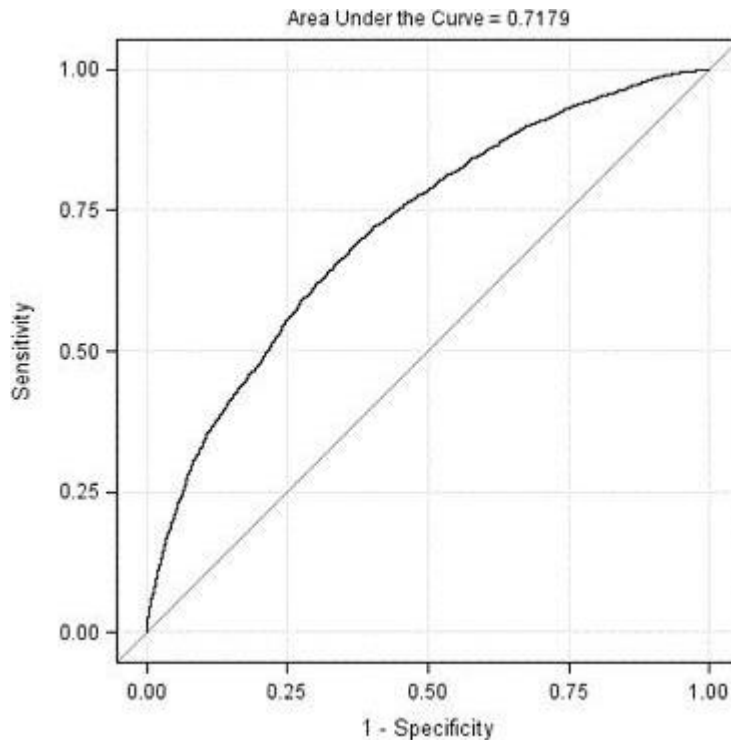
Note:- You can use calculator for this question

a) 7000
b) 9000
c) 7009
d) 9007

Answer: a

Question: 4

Refer to the exhibit:

Area Under the Curve = 0.7179

For the ROC curve shown, what is the meaning of the area under the curve?

a) percent concordant plus percent tied
b) percent concordant plus (.5 * percent tied)
c) percent concordant plus (.5 * percent discordant)
d) percent discordant plus percent tied

Answer: b

Question: 5

An analyst is screening for irrelevant variables by estimating strength of association between each input and the target variable. The analyst is using Spearman correlation and Hoeffding's D statistics in the CORR procedure.

What would likely cause some inputs to have a large Hoeffding and a near zero Spearman statistic?

a) nonmonotonic association between the variables
b) linear association between the variables
c) monotonic association between the variables
d) no association between the variables

Answer: a

Question: 6

Which statement is an assumption of logistic regression?

    a) The sample size is greater than 100.
    b) The logit is a linear function of the predictors.
    c) The predictor variables are not correlated.
    d) The errors are normally distributed.

Answer: b

Question: 7

A financial analyst wants to know whether assets in portfolio A are more risky (have higher variance) than those in portfolio B.

The analyst computes the annual returns (or percent changes) for assets within each of the two groups and obtains the following output from the GLM procedure:

Levene's Test for Homogeneity of Return Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Portfolio	1	217.2	217.2	16.29	0.0005
Error	23	306.7	13.3352		

Which conclusion is supported by the output?

    a) Assets in portfolio A are significantly more risky than assets in portfolio B.
    b) Assets in portfolio B are significantly more risky than assets in portfolio A.
    c) The portfolios differ significantly with respect to risk.
    d) The portfolios do not differ significantly with respect to risk.

Answer: c

Question: 8

When selecting variables or effects using SELECTION=BACKWARD in the LOGISTIC procedure, the business analyst's model selection terminated at Step 3.

What happened between Step 1 and Step 2?

    a) DF increased.
    b) AIC increased.
    c) Pr > Chisq increased.
    d) - 2 Log L increased.

Answer: d

Question: 9

Refer to the REG procedure output:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	618.44051	40.03665	15.45	<.0001	0
overhead	1	4.99845	0.00157	3181.24	<.0001	0.99993
scrap	1	2.82667	0.71581	3.95	<.0001	0.00124
training	1	-50.95436	2.82069	-18.06	<.0001	-0.00568

What is the most important predictor of the response variable?

  a) intercept
  b) overhead
  c) scrap
  d) raining

Answer: b

Question: 10

A linear model has the following characteristics:
- a dependent variable (y)
- one continuous predictor variables (x1) including a quadratic term (x12)
- one categorical predictor variable (c1 with 3 levels)
- one interaction term (c1 by x1)

Which SAS program fits this model?

  a) proc glm data=SASUSER.MLR;
     class c1;
     model y = c1 x1 x1sq c1byx1 /solution;
     run;
  b) proc reg data=SASUSER.MLR;
     model y = c1 x1 x1sq c1byx1 /solution;
     run;
  c) proc glm data=SASUSER.MLR;
     class c1;
     model y = c1 x1 x1*x1 c1*x1 /solution;
     run;
  d) proc reg data=SASUSER.MLR;
     model y = c1 x1 x1*x1 c1*x1;
     run;

Answer: c

Study Guide to Crack SAS Statistical Business Analyst A00-240 Exam:

- Getting details of the A00-240 syllabus, is the first step of a study plan. This pdf is going to be of ultimate help. Completion of the syllabus is must to pass the A00-240 exam.
- Making a schedule is vital. A structured method of preparation leads to success. A candidate must plan his schedule and follow it rigorously to attain success.
- Joining the SAS provided training for A00-240 exam could be of much help. If there is specific training for the exam, you can discover it from the link above.
- Read from the A00-240 sample questions to gain your idea about the actual exam questions. In this PDF useful sample questions are provided to make your exam preparation easy.
- Practicing on A00-240 practice tests is must. Continuous practice will make you an expert in all syllabus areas.

Reliable Online Practice Test for A00-240 Certification

Make AnalyticsExam.Com your best friend during your SAS Certified Statistical Business Analyst Using SAS 9 exam preparation. We provide authentic practice tests for the A00-240 exam. Experts design these online practice tests, so we can offer you an exclusive experience of taking the actual A00-240 exam. We guarantee you 100% success in your first exam attempt if you continue practicing regularly. Don't bother if you don't get 100% marks in initial practice exam attempts. Just utilize the result section to know your strengths and weaknesses and prepare according to that until you get 100% with our practice tests. Our evaluation makes you confident, and you can score high in the A00-240 exam.

Start Online Practice of A00-240 Exam by Visiting URL

https://www.analyticsexam.com/sas-certification/a00-240-sas-certified-statistical-business-analyst-using-sas-9-regression-and