A00-240 – SAS STATISTICAL BUSINESS ANALYST CERTIFICATION QUESTIONS AND STUDY GUIDE

SAS Certified Statistical Business Analyst Using SAS 9 (A00-240)

Contents

SAS Statistical Business Analyst Certification Details

Exam Name	SAS Certified Statistical Business Analyst Using SAS 9 - Regression and Modeling
Exam Code	A00-240
Duration	120 Minutes
Passing Percentage	68%
Negative Marking	No Negative Marking
Partial Credit	No
Reference Book	
Training	Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression Predictive Modeling Using Logistic Regression
Schedule Your exam	Pearson VUE
Sample Questions	SAS Statistical Business Analyst Certification Sample Question
Recommended Practice tool	**SAS Statistical Business Analyst Certification Practice Exam**

SAS Statistical Business Analyst Certification Syllabus for A00-240 (Study Aid)

ANOVA (10%)

- Verify the assumptions of ANOVA
- Analyze differences between population means using the GLM and TTEST procedures
- Perform ANOVA post hoc test to evaluate treatment effect
- Detect and analyze interactions between factors

Linear Regression (20%)

- Fit a multiple linear regression model using the REG and GLM procedures
- Analyze the output of the REG procedure for multiple linear regression models
- Use the REG procedure to perform model selection
- Assess the validity of a given regression model through the use of diagnostic and residual analysis

Logistic Regression (25%)

- Perform logistic regression with the LOGISTIC procedure
- Optimize model performance through input selection
- Interpret the output of the LOGISTIC procedure
- Score new data sets using the LOGISTIC and SCORE procedures

Prepare Inputs for Predictive Model Performance (20%)

- Identify potential problems with input data
- Use the DATA step to manipulate data with loops, arrays, conditional statements and functions
- Reduce the number of categorical levels in a predictive model
- Screen variables for irrelevance using the CORR procedure
- Screen variables for non-linearity using empirical logit plots

Measure Model Performance (25%)

- Apply the principles of honest assessment to model performance measurement
- Assess classifier performance using the confusion matrix
- Model selection and validation using training and validation data
- Create and interpret graphs (ROC, lift, and gains charts) for model comparison and selection

Establish effective decision cut-off values for scoring

SAS Statistical Business Analyst Exam (A00-240) Sample Questions

- Below are the 10 sample questions which will help you be familiar with SAS Certified Statistical Business Analyst Using SAS 9 - Regression and Modeling (A00-240) exam style and Structure.

- These questions are just for demonstration purpose, there are many scenario based question are included in **Premium SAS Statistical Business Analyst Practice Exam**

- Access to all 80+ questions is available only through premium practice exam available to members at www.analyticsexam .com

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;

QUESTION 2: 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;
  E.  proc reg data=SASUSER.MLR;
            model y = c1 x1 x1*x1 c1*x1;
        run;
```

QUESTION 3:

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.

QUESTION 4: 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. training

QUESTION 5: 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.

QUESTION 6: 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.

QUESTION 7: 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?

A. 7000
B. 6000
C. 10000
D. 1000

QUESTION 8: 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

QUESTION 9: 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.

QUESTION 10: Refer to the exhibit:

Area Under the Curve = 0.7179
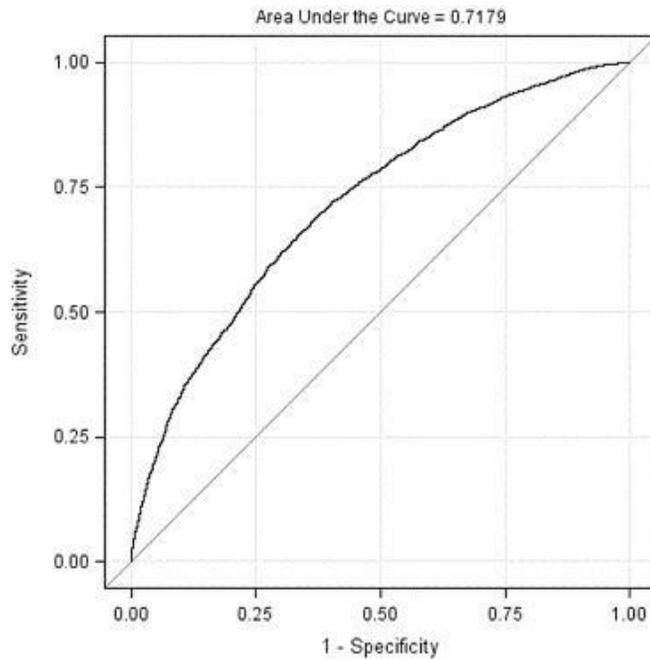
For the ROC curve shown, what is the meaning of the area under the curve?

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

Answers:

Question: 1	Answer: C	Question: 2	Answer: C
Question: 3	Answer: B	Question: 4	Answer: B
Question: 5	Answer: B	Question: 6	Answer: D
Question: 7	Answer: A	Question: 8	Answer: A
Question: 9	Answer: D	Question: 10	Answer: B