

Exam Questions A00-240

SAS Certified Statistical Business Analyst Using SAS 9: Regression and Modeling Credential

<https://www.2passeasy.com/dumps/A00-240/>



NEW QUESTION 1

Select the equivalent LOGISTIC procedure model statements. (Choose two.)

- A. Mode1 Purchase * Gender Age Region;
- B. Mode1 Purchase * Gender | Age | Region;
- C. Mode1 Purchase * Gender|Age|Region @1;
- D. Mode1 Purchase * Gender|Age|Region @2;

Answer: AC

NEW QUESTION 2

Identify the correct SAS program for fitting a multiple linear regression model with dependent variable (y) and four predictor variables (x1-x4).

- A.

```
proc reg data=SASUSER.MLR;
  var y x1 x2 x3 x4;
  model y = x1-x4;
run;
```
- B.

```
proc reg data=SASUSER.MLR;
  model y = x1-x4;
run;
```
- C.

```
proc reg data=SASUSER.MLR;
  model y = x1;
  model y = x2;
  model y = x3;
  model y = x4;
run;
```
- D.

```
proc reg data=SASUSER.MLR;
  model y = x1 x2 x3 x4 /solution;
run;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: B

NEW QUESTION 3

Refer to the REG procedure output:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	31848	15924	13.42	<.0001
Error	97	115082	1186.40833		
Corrected Total	99	146930			

Root MSE	34.44428	R-Square	0.2168
Dependent Mean	606.38715	Adj R-Sq	0.2006
Coeff Var	5.68025		

An analyst has selected this model as a champion because it shows better model fit than a competing model with more predictors. Which statistic justifies this rationale?

- A. R-Square
- B. Coeff Var
- C. Adj R-Sq
- D. Error DF

Answer: C

NEW QUESTION 4

What is the default method in the LOGISTIC procedure to handle observations with missing data?

- A. Missing values are imputed.
- B. Parameters are estimated accounting for the missing values.
- C. Parameter estimates are made on all available data.
- D. Only cases with variables that are fully populated are used.

Answer: D

NEW QUESTION 5

Consider scoring new observations in the SCORE procedure versus the SCORE statement in the LOGISTIC procedure. Which statement is true?

- A. The SCORE statement in the LOGISTIC procedure returns only predicted probabilities, whereas the SCORE procedure returns only predicted logits.
- B. The SCORE statement in the LOGISTIC procedure returns only predicted logits, whereas the SCORE procedure returns only predicted probabilities.
- C. Unlike the SCORE procedure, the SCORE statement in the LOGISTIC procedure produces both predicted probabilities and predicted logits.
- D. The SCORE procedure and the SCORE statement in the LOGISTIC procedure produce the same output.

Answer: A

NEW QUESTION 6

Which of the following describes a concordant pair of observations in the LOGISTIC procedure?

- A. An observation with the event has an equal probability as another observation with the event.
- B. An observation with the event has a lower predicted probability than the observation without the event.
- C. An observation with the event has an equal predicted probability as the observation without the event.
- D. An observation with the event has a higher predicted probability than the observation without the event

Answer: D

NEW QUESTION 7

Given the following LOGISTIC procedure:

```
proc logistic data = MYDIR.CONVERT des outest=OUTFILE_1;  
  model Attrite = Calls Plan Billing_code;  
  score data=MYDIR.NEW_ATTRITE_DATA out=OUTFILE_2;  
run;
```

What is the difference between the datasets OUTFILEJ and OUTFILE_2?

- A. OUTFILE_1 contains the final parameter estimates while OUTFILE_2 contains the newly scored probabilities.
- B. OUTFILE_1 contains the model goodness of fit statistics while OUTFILE_2 contains the newly scored probabilities
- C. OUTFILE_1 contains the model goodness of fit statistics while OUTFILE_2 contains the newly scored logits.
- D. OUTFILEJ contains the final parameter estimates and Wald Chi-Square values while OUTFILE_2 contains the newly scored probabilities.

Answer: A

NEW QUESTION 8

Suppose training data are oversampled in the event group to make the number of events and non-events roughly equal. A logistic regression is run and the probabilities are output to a data set NEW and given the variable name PE. A decision rule considered is, "Classify data as an event if probability is greater than 0.5." Also the data set NEW contains a variable TG that indicates whether there is an event (1=Event, 0= No event). The following SAS program was used.

```
data NEW;  
  set NEW;  
  Solicit = PE > .5;  
run;  
proc means data=NEW(where = (TG=1)) mean;  
  var Solicit;  
run;
```

What does this program calculate?

- A. Depth
- B. Sensitivity
- C. Specificity
- D. Positive predictive value

Answer: B

NEW QUESTION 9

A company has branch offices in eight regions. Customers within each region are classified as either "High Value" or "Medium Value" and are coded using the variable name VALUE. In the last year, the total amount of purchases per customer is used as the response variable. Suppose there is a significant interaction between REGION and VALUE. What can you conclude?

- A. More high value customers are found in some regions than others.
- B. The difference between average purchases for medium and high value customers depends on the region.
- C. Regions with higher average purchases have more high value customers.
- D. Regions with higher average purchases have more medium value customers.

Answer: B

NEW QUESTION 10

An analyst fits a logistic regression model to predict whether or not a client will default on a loan. One of the predictors in the model is agent, and each agent serves 15-20 clients each. The model fails to converge. The analyst prints the summarized data, showing the number of defaulted loans per agent. See the partial output below:

Obs	agent	clients	defaults
1	1	17	12
2	2	19	0
3	3	16	7
4	4	15	5
5	5	19	13
6	6	17	8
7	7	16	9
8	8	17	10
9	9	17	11
10	10	16	8

What is the most likely reason that the model fails to converge?

- A. There is quasi-complete separation in the data.
- B. There is collinearity among the predictors.
- C. There are missing values in the data.
- D. There are too many observations in the data.

Answer: A

NEW QUESTION 10

Refer to the exhibit:

Number in Model	R-Square	Adjusted R-Square	C(p)	AIC	Root MSE	SBC	Variables in Model
1	0.7434	0.7345	13.6988	64.5341	2.74478	67.40210	RunTime
1	0.1595	0.1305	106.3021	101.3131	4.96748	104.18108	RestPulse
2	0.7642	0.7474	12.3894	63.9050	2.67739	68.20695	Age RunTime
2	0.7614	0.7444	12.8372	64.2740	2.69337	68.57597	RunTime RunPulse
3	0.8111	0.7901	6.9596	59.0373	2.44063	64.77326	Age RunTime RunPulse
3	0.8100	0.7889	7.1350	59.2183	2.44777	64.95424	RunTime RunPulse MaxPulse
4	0.8368	0.8117	4.8800	56.4995	2.31159	63.66941	Age RunTime RunPulse MaxPulse
4	0.8165	0.7883	8.1035	60.1386	2.45133	67.30850	Age Weight RunTime RunPulse
5	0.8480	0.8176	5.1063	56.2986	2.27516	64.90250	Age Weight RunTime RunPulse MaxPulse
5	0.8370	0.8044	6.8461	58.4590	2.35583	67.06288	Age RunTime RunPulse RestPulse MaxPulse
6	0.8487	0.8108	7.0000	58.1616	2.31695	68.19952	Age Weight RunTime RunPulse RestPulse MaxPulse

SAS output from the RSQUARE selection method, within the REG procedure, is shown. The top two models in each subset are given. Based on the AIC statistic, which model is the champion model?

- A. Age Weight RunTime RunPulse MaxPulse
- B. Age Weight RunTime RunPulse RestPulse MaxPulse
- C. RestPulse
- D. RunTime

Answer: A

NEW QUESTION 15

Assume a \$10 cost for soliciting a non-responder and a \$200 profit for soliciting a responder. The logistic regression model gives a probability score named P_R on a SAS data set called VALID. The VALID data set contains the responder variable Pinch, a 1/0 variable coded as 1 for responder. Customers will be solicited when their probability score is more than 0.05.

Which SAS program computes the profit for each customer in the data set VALID?

- A.

```
data VALID;
  set VALID;
  Profit = (P_R > .05)*Purch*200 - (P_R > .05)*(1 - Purch)*10;
run;
```
- B.

```
data VALID;
  set VALID;
  Profit = (P_R <= .05)*Purch*200 - (P_R > .05)*(1 - Purch)*10;
run;
```
- C.

```
data VALID;
  set VALID;
  if P_R > .05;
  Profit = (P_R > .05)*Purch*200 - (P_R > .05)*(1 - Purch)*10;
run;
```
- D.

```
data VALID;
  set VALID;
  if P_R >.05;
  Profit = (P_R > .05)*Purch*200 + (P_R <= .05)*(1 - Purch)*10;
run;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: A

NEW QUESTION 20

The total modeling data has been split into training, validation, and test data. What is the best data to use for model assessment?

- A. Training data
- B. Total data
- C. Test data
- D. Validation data

Answer: D

NEW QUESTION 25

The following LOGISTIC procedure output analyzes the relationship between a binary response and an ordinal predictor variable, wrist_size Using reference cell coding, the analyst selects Large (L) as the reference level.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.0415	0.4749	4.8101	0.0283
wrist_size M	1	1.1234	0.4989	5.0697	0.0243
wrist_size S	1	1.6078	0.5478	8.6133	0.0033

What is the estimated logit for a person with large wrist size? Click the calculator button to display a calculator if needed. A. 0.0819

- A. 0.5663
- B. -3.7727
- C. -1.0415

Answer:

NEW QUESTION 29

Given the following SAS data set TEST:

```
Inc_Group
 1
 2
 3
 4
 5
```

Which SAS program is NOT a correct way to create dummy variables?

- A.

```
data DUMMY_TEST1;
  set TEST;
  Inc_Group1=(Inc_Group=1);
  Inc_Group2=(Inc_Group=2);
  Inc_Group3=(Inc_Group=3);
  Inc_Group4=(Inc_Group=4);
  Inc_Group5=(Inc_Group=5);
run;
```
- B.

```
data DUMMY_TEST1;
  set TEST;
  if Inc_Group=1 then Inc_Group1=1;
  else Inc_Group1=0;
  if Inc_Group=2 then Inc_Group2=1;
  else Inc_Group2=0;
  if Inc_Group=3 then Inc_Group3=1;
  else Inc_Group3=0;
  if Inc_Group=4 then Inc_Group4=1;
  else Inc_Group4=0;
  if Inc_Group=5 then Inc_Group5=1;
  else Inc_Group5=0;
run;
```
- C.

```
data DUMMY_TEST1 (drop=i);
  set TEST;
  array inc(*) Inc_Group1 - Inc_Group5;
  do i = 1 to 5;
    inc(i) = ( Inc_Group = i );
  end;
run;
```
- D.

```
data DUMMY_TEST1 (drop=i);
  set TEST;
  array inc(*) Inc_Group1 Inc_Group2 Inc_Group3
             Inc_Group4 Inc_Group5;
  do i = 1 to 5;
    ( Inc_Group = i );
  end;
run;
```

- A. Option A
 B. Option B
 C. Option C
 D. Option D

Answer: D

NEW QUESTION 34

This question will ask you to provide a missing option. Given the following SAS program:

```
proc corr data = MYDATA <insert option here> ;
  var x1 x2 x3 x4 x5;
  with Target;
run;
```

What option must be added to the program to obtain a data set containing Pearson statistics?

- A. OUTPUT=estimates
 B. OUTP=estimates
 C. OUTSTAT=estimates
 D. OUTCORR=estimates

Answer: B

NEW QUESTION 37

There are missing values in the input variables for a regression application.

Which SAS procedure provides a viable solution?

- A. GLM
 B. VARCLUS
 C. STDI2E
 D. CLUSTER

Answer: C

NEW QUESTION 40

Refer to the REG procedure output:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	33033	11011	115.63	<.0001
Error	496	47231	95.22454		
Corrected Total	499	80265			

Click on the calculator button to display a calculator if needed.

- A. 0.4115
- B. 0.6994
- C. 0.5884
- D. 0.1372

Answer: A

NEW QUESTION 42

Which SAS program will detect collinearity in a multiple regression application?

- A.

```
proc reg data = SASUSER.RETAIL;
  model Purchase = Gender Age Income / lackfit;
run;
```
- B.

```
proc reg data = SASUSER.RETAIL;
  model Purchase = Gender Age Income / vif;
run;
```
- C.

```
proc reg data=SASUSER.RETAIL plots(only)=(COOKSD);
  model Purchase = Gender Age Income;
run;
```
- D.

```
proc reg data=sasuser.retail plots(only)=(RSTUDENTBYPREDICTED);
  model Purchase = Gender Age Income;
run;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: B

NEW QUESTION 43

The standard form of a linear regression model is:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Which statement best summarizes the assumptions placed on the errors?

- A. The errors are correlated, normally distributed with constant mean and zero variance.
- B. The errors are correlated, normally distributed with zero mean and constant variance.
- C. The errors are independent, normally distributed with constant mean and zero variance.
- D. The errors are independent, normally distributed with zero mean and constant variance.

Answer: D

NEW QUESTION 44

The question will ask you to provide a missing statement. Given the following SAS program:

```
proc logistic data = MYDIR.DEFAULT_DATA des;
  model Purchase = Money Acct_type Debt Employment;
  <insert statement here>
run;
```

Which SAS statement will complete the program to correctly score the data set NEW_DATA?

- A. Scoredata data=MYDIR.NEW_DATA out=scores;
- B. Scoredata data=MYDIR.NEW_DATA output=scores;
- C. Scoredata=HYDIR.NEU_DATA output=scores;
- D. Scoredata=MYDIR,NEW DATA out=scores;

Answer: D

NEW QUESTION 46

A financial services manager wants to assess the probability that certain clients will default on their Home Equity Line of Credit (HELOC). A former employee left the code listed below.

```
proc logistic data = MYDIR.HELOC des outest=MSG;
  model DEFAULT = amount job_code years_at_residence;
run;

proc score data = MYDIR.RECENT_HELOC
  out = SCORED_HELOC
  score = MSG
  type = parms;
  var Amount Job_code Years_at_residence;
run;
```

The training data set is named HELOC, while a similar data set of more recent clients is named RECENT_HELOC. Which SAS data steps will calculate the predicted probability of default on recent clients? (Choose two.)

- A.

```
data NEW_PROB;
  set SCORED_HELOC;
  p=1/(1+exp(-DEFAULT));
run;
```
- B.

```
data NEW_PROB;
  set SCORED_HELOC;
  ODDS = exp(DEFAULT);
  p = ODDS / (1+ODDS);
run;
```
- C.

```
data NEW_PROB;
  set SCORED_HELOC;
  p=(1+exp(DEFAULT))/exp(DEFAULT);
run;
```
- D.

```
data NEW_PROB;
  set SCORED_HELOC;
  p = DEFAULT / (1+DEFAULT);
run;
```

- A. Option A
B. Option B
C. Option C
D. Option D

Answer: AB

NEW QUESTION 50

What is a drawback to performing data cleansing (imputation, transformations, etc.) on raw data prior to partitioning the data for honest assessment as opposed to performing the data cleansing after partitioning the data?

- A. It violates assumptions of the model.
B. It requires extra computational effort and time.
C. It omits the training (and test) data sets from the benefits of the cleansing methods.
D. There is no ability to compare the effectiveness of different cleansing methods.

Answer: D

NEW QUESTION 54

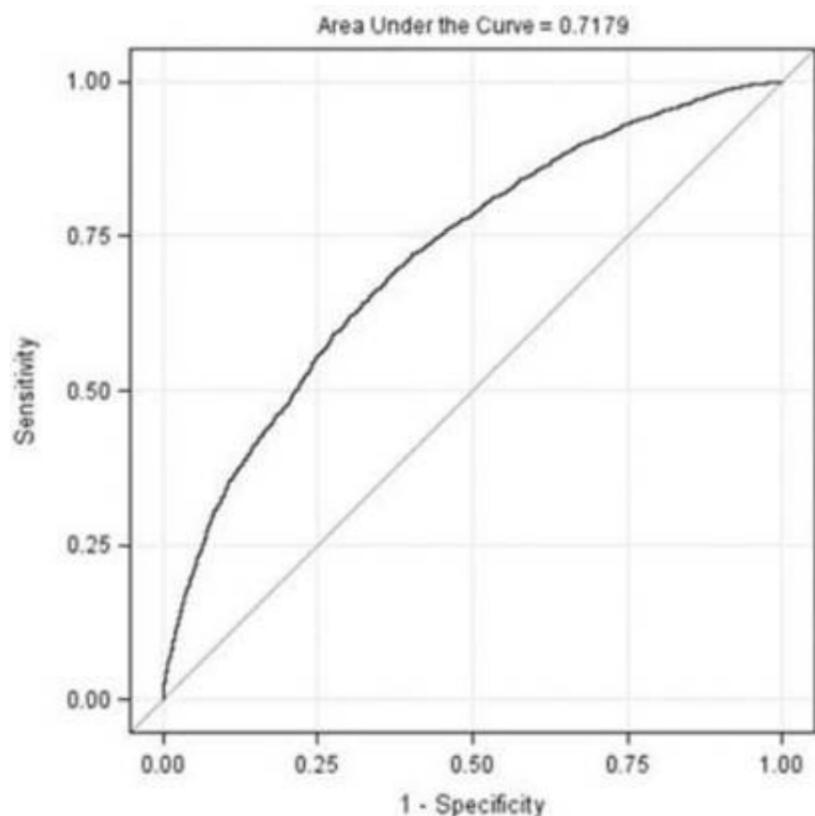
Which method is NOT an appropriate way to score new observations with a known target in a logistic regression model?

- A. Use the SCORE statement in the LOGISTIC procedure.
B. Augment the training data set with new observations and set their responses to missing.
C. Augment the training data set with new observations and rerun the LOGISTIC procedure.
D. Use the saved parameter estimates from the LOGISTIC procedure and score new observations in the SCORE procedure.

Answer: C

NEW QUESTION 59

Refer to the ROC curve:



As you move along the curve, what changes?

- A. The priors in the population
- B. The true negative rate in the population
- C. The proportion of events in the training data
- D. The probability cutoff for scoring

Answer: D

NEW QUESTION 63

Which statistic, calculated from a validation sample, can help decide which model to use for prediction of a binary target variable?

- A. Adjusted R Square
- B. Mallows's Cp
- C. Chi Square
- D. Average Squared Error

Answer: D

NEW QUESTION 64

Which SAS program will divide the original data set into 60% training and 40% validation data sets, stratified by county?

- A.

```
proc surveyselect data=SASUSER.DATABASE samprate=0.6 out=sample;
  strata county;
run;
```
- B.

```
proc sort data=SASUSER.DATABASE;
  by county;
run;
proc surveyselect data=SASUSER.DATABASE samprate=0.6 out=sample outall;
run;
```
- C.

```
proc sort data=SASUSER.DATABASE;
  by county;
run;
proc surveyselect data=SASUSER.DATABASE samprate =0.6 out=sample outall;
  strata county;
run;
```
- D.

```
proc sort data=SASUSER.DATABASE;
  by county;
run;
proc surveyselect data=SASUSER.DATABASE samprate =0.6 out=sample;
  strata county;
run;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: C

NEW QUESTION 68

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

The Intercept estimate is interpreted as:

- A. The predicted value of the response when all the predictors are at their current values.
- B. The predicted value of the response when all predictors are at their means.
- C. The predicted value of the response when all predictors = 0.
- D. The predicted value of the response when all predictors are at their minimum values.

Answer: C

NEW QUESTION 73

Refer to the confusion matrix:

		Predicted Outcome	
		0	1
Actual Outcome	0	58	44
	1	23	25

Calculate the sensitivity. (0 - negative outcome, 1 - positive outcome) Click the calculator button to display a calculator if needed.

- A. 25/48
- B. 58/102
- C. 25/B9
- D. 58/81

Answer: A

NEW QUESTION 77

Refer to the following odds ratio table:

Odds Ratio Estimates and Profile-Likelihood Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
salary	1.0000	1.142	1.083	1.220

What is a correct interpretation of the estimate?

- A. The odds of the event are 1.142 greater for each one dollar increase in salary.
- B. The odds of the event are 1.142 greater for each one thousand dollar increase in salary.
- C. The probability of the event is 1.142 greater for each one dollar increase in salary.
- D. The probability of the event is 1.142 greater for each one thousand dollar increase in salary.

Answer: B

NEW QUESTION 80

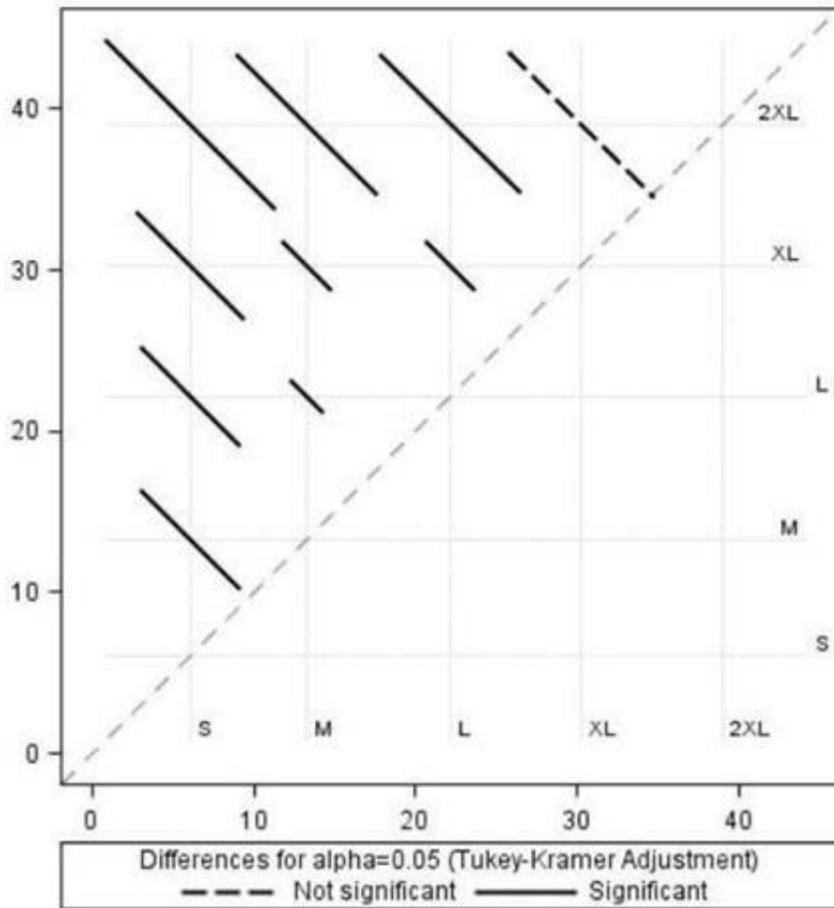
A predictive model uses a data set that has several variables with missing values. What two problems can arise with this model? (Choose two.)

- A. The model will likely be overfit.
- B. There will be a high rate of collinearity among input variables.
- C. Complete case analysis means that fewer observations will be used in the model building process.
- D. New cases with missing values on input variables cannot be scored without extra data processing.

Answer: CD

NEW QUESTION 82

Refer to the exhibit.



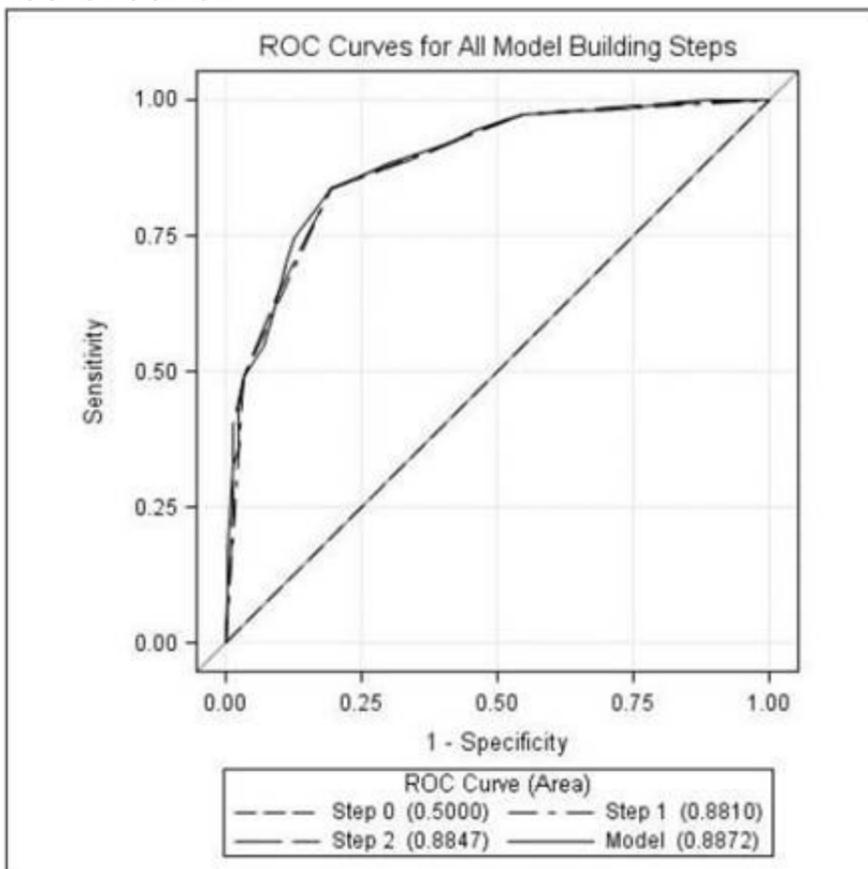
Based on the control plot, which conclusion is justified regarding the means of the response?

- A. All groups are significantly different from each other.
- B. 2XL is significantly different from all other groups.
- C. Only XL and 2XL are not significantly different from each other.
- D. No groups are significantly different from each other.

Answer: C

NEW QUESTION 84

Refer to the exhibit:



An analyst examined logistic regression models for predicting whether a customer would make a purchase. The ROC curve displayed summarizes the models. Using the selected model and the analyst's decision rule, 25% of the customers who did not make a purchase are incorrectly classified as purchasers. What can be concluded from the graph?

- A. About 25% of the customers who did make a purchase are correctly classified as making a purchase.
- B. About 50% of the customers who did make a purchase are correctly classified as making a purchase.
- C. About 85% of the customers who did make a purchase are correctly classified as making a purchase.
- D. About 95% of the customers who did make a purchase are correctly classified as making a purchase.

Answer: C

NEW QUESTION 85

An analyst generates a model using the LOGISTIC procedure. They are now interested in getting the sensitivity and specificity statistics on a validation data set for a variety of cutoff values.

Which statement and option combination will generate these statistics?

- A. Scoredata=valid1 out=roc;
- B. Scoredata=valid1 outroc=roc;
- C. mode1resp(event= '1') = gender region/outroc=roc;
- D. mode1resp(event="1") = gender region/ out=roc;

Answer: B

NEW QUESTION 89

Customers were surveyed to assess their intent to purchase a product. An analyst divided the customers into groups defined by the company's pre-assigned market segments and tested for difference in the customers' average intent to purchase. The following is the output from the GLM procedure:

Source	DF	Sum of		F Value	Pr > F
		Squares	Mean Square		
Model	7	15716.87902	2245.26843	64.98	<.0001
Error	146	5044.56579	34.55182		
Corrected Total	153	20761.44481			

What percentage of customers' intent to purchase is explained by market segment? Click the calculator button to display a calculator if needed.

- A. <0.01%
- B. 35%
- C. 65%
- D. 76%

Answer: D

NEW QUESTION 92

An analyst has a sufficient volume of data to perform a 3-way partition of the data into training, validation, and test sets to perform honest assessment during the model building process.

What is the purpose of the test data set?

- A. To provide a unbiased measure of assessment for the final model.
- B. To compare models and select and fine-tune the final model.
- C. To reduce total sample size to make computations more efficient.
- D. To build the predictive models.

Answer: A

NEW QUESTION 94

A marketing manager attempts to determine those customers most likely to purchase additional products as the result of a nation-wide marketing campaign. The manager possesses a historical dataset (CAMPAIGN) of a similar campaign from last year.

It has the following characteristics:

- ? Target variable Respond (0,1)
- ? Continuous predictor Income
- ? Categorical predictor Homeowner(Y,N)

Which SAS program performs this analysis?

- A.

```
proc logistic data=MYDIR.CAMPAIGN descending;
  class Homeowner;
  model Respond = Income Homeowner;
run;
```
- B.

```
proc logistic data = MYDIR.CAMPAIGN descending;
  by Homeowner;
  model Respond = Income Homeowner;
run;
```
- C.

```
proc logistic data = MYDIR.CAMPAIGN descending;
  model Respond = Income Homeowner;
run;
```
- D.

```
proc logistic data = MYDIR.CAMPAIGN descending;
  class Income Homeowner;
  model Respond = Income Homeowner;
run;
```

- A. Option A
- B. Option B
- C. Option C
- D. Option D

Answer: A

NEW QUESTION 97

CORRECT TEXT

A linear model has the following characteristics:

*A dependent variable (y)

*One continuous variable (x1), including a quadratic term (x1²)

*One categorical (d with 3 levels) predictor variable and an interaction term (d by x1) How many parameters, including the intercept, are associated with this model?

Enter your numeric answer in the space below. Do not add leading or trailing spaces to your answer.

Answer:

Explanation: 7

NEW QUESTION 100

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual A00-240 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the A00-240 Product From:

<https://www.2passeasy.com/dumps/A00-240/>

Money Back Guarantee

A00-240 Practice Exam Features:

- * A00-240 Questions and Answers Updated Frequently
- * A00-240 Practice Questions Verified by Expert Senior Certified Staff
- * A00-240 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * A00-240 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year