

100% Money Back
Guarantee

Vendor:SASInstitute

Exam Code:A00-240

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

Version:Demo

QUESTION 1

A marketing campaign will send brochures describing an expensive product to a set of customers. The cost for mailing and production per customer is \$50. The company makes \$500 revenue for each sale. What is the profit matrix for a typical person in the population?

A.

	Purchase	
Solicit	No	Yes
No	-50	0
Yes	0	450

B.

	Purchase	
Solicit	No	Yes
No	0	0
Yes	-50	500

C.

	Purchase	
Solicit	No	Yes
No	0	0
Yes	-50	450

D.

	Purchase	
Solicit	No	Yes
No	-50	0
Yes	0	500

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: C

QUESTION 2

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.

Correct Answer: D

QUESTION 3

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

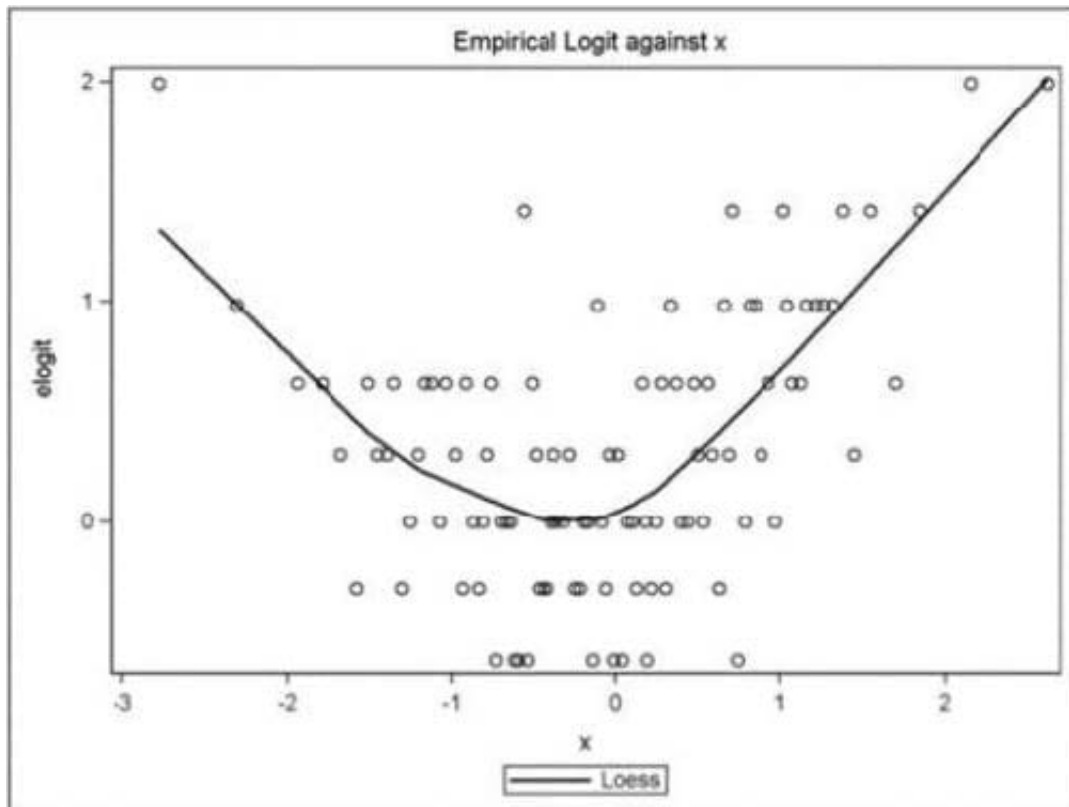
C. Option C

D. Option D

Correct Answer: AB

QUESTION 4

Refer to the following exhibit:



What is a correct interpretation of this graph?

- A. The association between the continuous predictor and the binary response is quadratic.
- B. The association between the continuous predictor and the log-odds is quadratic.
- C. The association between the continuous predictor and the continuous response is quadratic.
- D. The association between the binary predictor and the log-odds is quadratic.

Correct Answer: B

QUESTION 5

What does the Pearson product moment correlation coefficient measure?

- A. nonlinear and nonmonotonic association between two variables
- B. linear and monotonic association between two variables
- C. linear and nonmonotonic association between two variables
- D. nonlinear and monotonic association between two variables

Correct Answer: B

Reference: http://d-scholarship.pitt.edu/8056/1/Chokns_etd2010.pdf

QUESTION 6

The standard form of a linear regression model is:

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

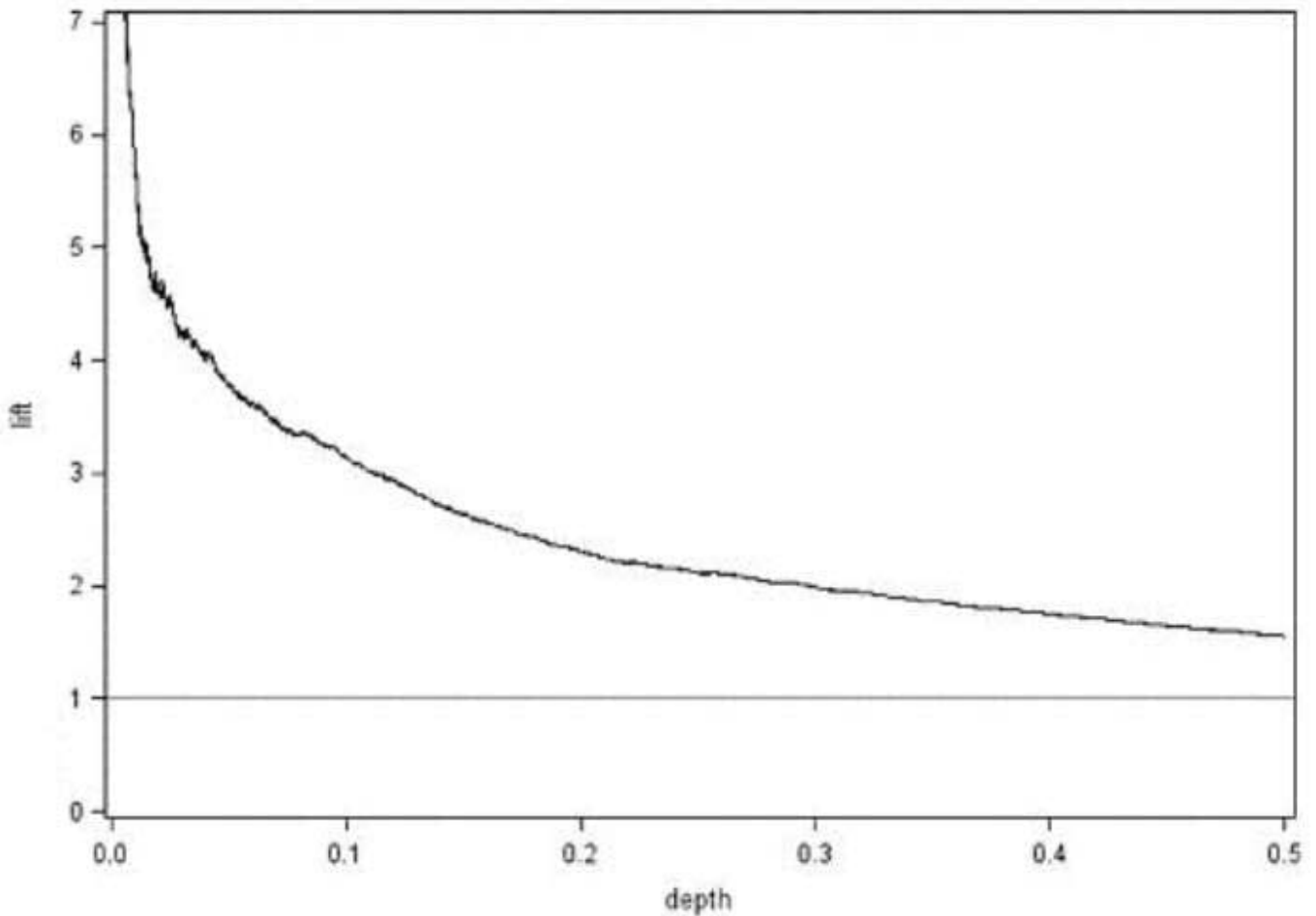
Which statement best summarizes the assumptions placed on the errors?

- 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.

Correct Answer: D

QUESTION 7

Refer to the lift chart:



At a depth of 0.1, Lift = 3.14. What does this mean?

- A. Selecting the top 10% of the population scored by the model should result in 3.14 times more events than a random draw of 10%.
- B. Selecting the observations with a response probability of at least 10% should result in 3.14 times more events than a random draw of 10%.
- C. Selecting the top 10% of the population scored by the model should result in 3.14 times greater accuracy than a random draw of 10%.
- D. Selecting the observations with a response probability of at least 10% should result in 3.14 times greater accuracy than a random draw of 10%.

Correct Answer: A

QUESTION 8

A non-contributing predictor variable ($\text{Pr} > |t| = 0.658$) is removed from an existing multiple linear regression model. What will be the result?

- A. An increase in R-Square
- B. A decrease in R-Square

- C. A decrease in Mean Square Error
- D. No change in R-Square

Correct Answer: B

QUESTION 9

This question will ask you to provide a missing option.

A business analyst is investigating the differences in sales figures across 8 sales regions. The analyst is interested in viewing the regression equation parameter estimates for each of the design variables.

Which option completes the program to produce the regression equation parameter estimates?

```
proc glm data=sales_spring;
    class region;
    model finalsales = region / <insert option here> ;
run;
quit;
```

- A. Solve
- B. Estimate
- C. Solution
- D. Est

Correct Answer: C

Reference: https://documentation.sas.com/?docsetId=statuganddocsetTarget=statug_ods_examples06.htm&docsetVersion=14.3&locale=en

QUESTION 10

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 OUTFILE_1 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.

Correct Answer: A

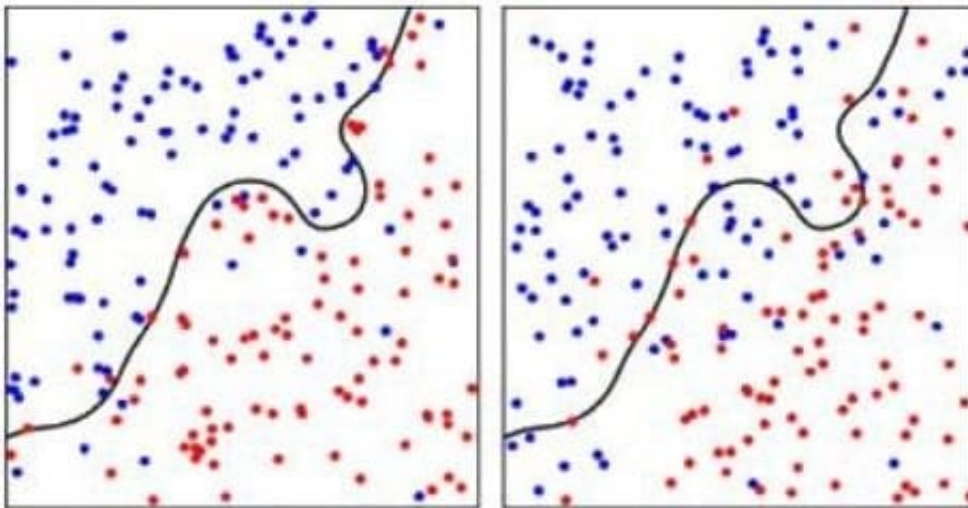
QUESTION 11

Refer to the exhibit:

Model A

training data

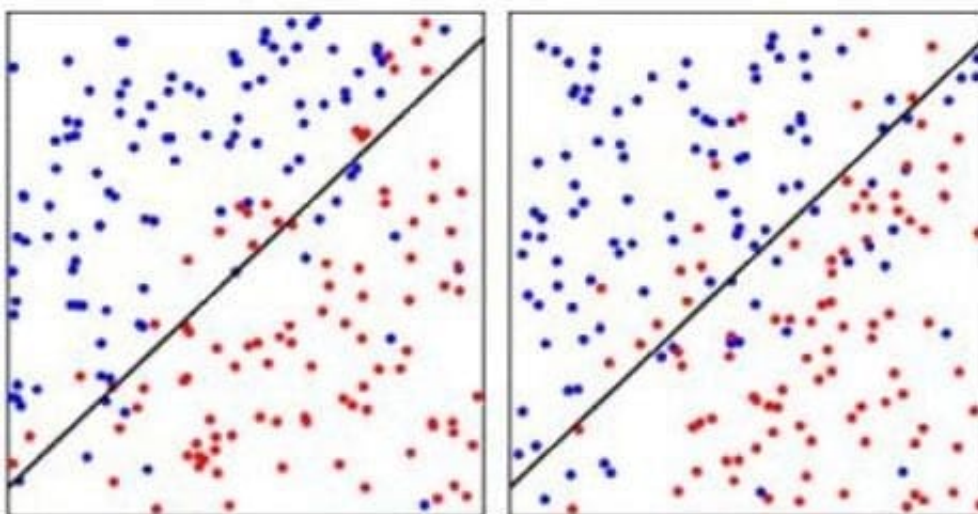
validation data



Model B

training data

validation data



The plots represent two models, A and B, being fit to the same two data sets, training and validation.

Model A is 90.5% accurate at distinguishing blue from red on the training data and 75.5% accurate at doing the same on validation data. Model B is 83% accurate at distinguishing blue from red on the training data and 78.3% accurate at doing the same on the validation data.

Which of the two models should be selected and why?

- A. Model A. It is more complex with a higher accuracy than model B on training data.
- B. Model A. It performs better on the boundary for the training data.
- C. Model B. It is more complex with a higher accuracy than model A on validation data.
- D. Model B. It is simpler with a higher accuracy than model A on validation data.

Correct Answer: D

QUESTION 12

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

Correct Answer: B

