

Looking for Real Exam Questions for IT Certification Exams!

We guarantee you can pass any IT certification exam at your first attempt with just 10-12 hours study of our guides.

Our study guides contain actual exam questions, you will get word to word same on your actual test; accurate answers with detailed explanation verified by experts and all graphics and drag-n-drop exhibits shown just as on the real test.

To test the quality of our guides, you can download the one-fourth portion of any guide from <http://www.certificationking.com> absolutely free.

Besides, we also offer complete version of following exams absolutely free. You can start your certification from these free guides and if you are satisfied you can buy the rest

- ◆ Microsoft: 70-270, 70-305 ◆ Cisco: 642-901 ◆ Oracle: 1Z0-007, 200 ◆ CompTIA: 220-601
- ◆ SUN: 310-014, 310-044 ◆ Citrix: 1Y0-A01, 1Y0-256 ◆ CIW: 1D0-420 ◆ Novell: 50-686
- ◆ Adobe: 9A0-029 ◆ Apple: 9L0-005, 9L0-505 ◆ Avaya: 132-S-100 ◆ Cognos: COG-105
- ◆ CWNP: PW0-100 ◆ EMC: E20-001 ◆ Hyperion: 4H0-002 ◆ HP: HP0-771, HP0-J24
- ◆ IBM: 000-253, 000-700 ◆ Juniper: JN0-100, JN0-201 ◆ Lotus: LOT-739
- ◆ Nortel: 920-803 ◆ SAS: A00-201 ◆ SNIA: S10-100 ◆ Sybase: 510-015
- ◆ Symantec: 250-101 ◆ TeraData: NR0-011

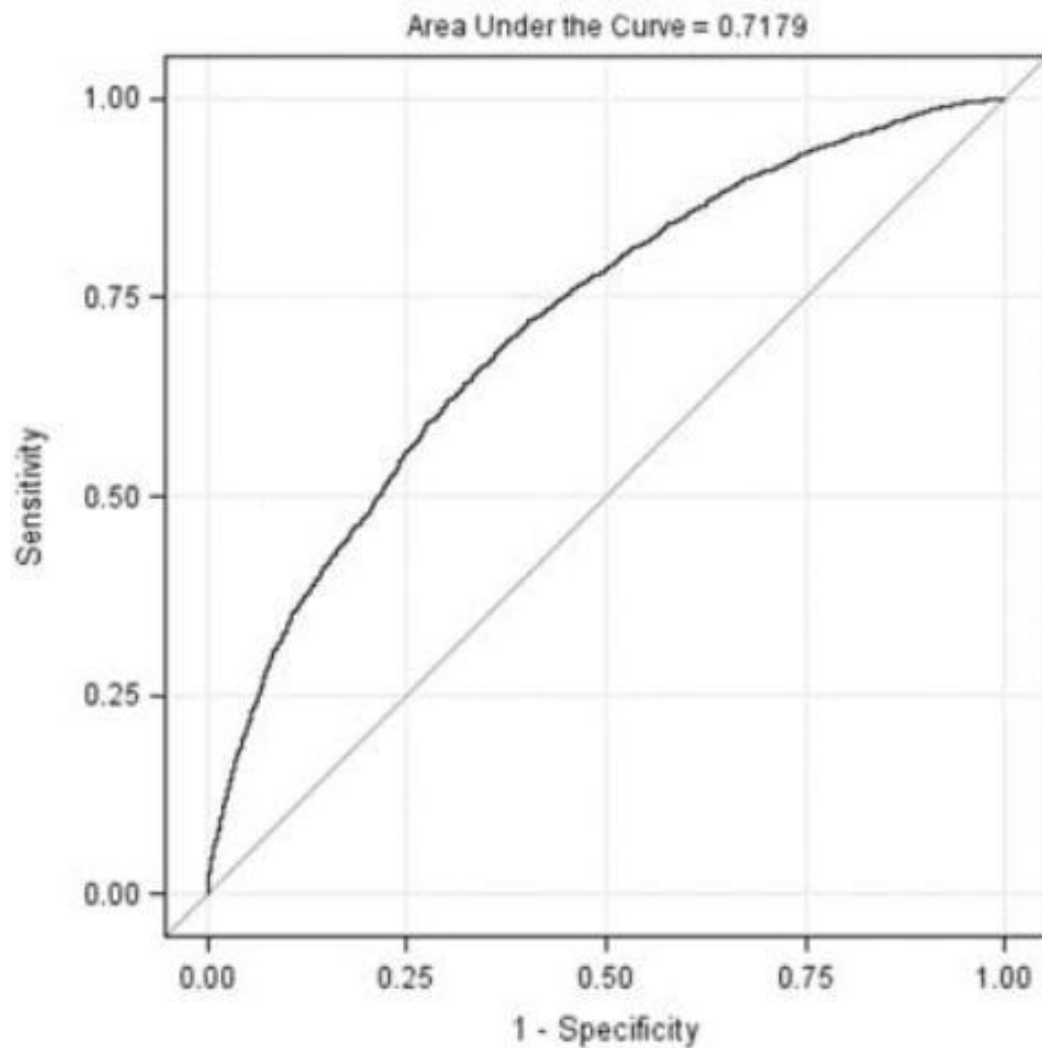
For pricing and placing order, please visit <http://certificationking.com/order.html>

We accept all major credit cards through www.paypal.com

For other payment options and any further query, feel free to mail us at info@certificationking.com

QUESTION NO: 1

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

Explanation:

QUESTION NO: 2

When mean imputation is performed on data after the data is partitioned for honest assessment,

what is the most appropriate method for handling the mean imputation?

- A. The sample means from the validation data set are applied to the training and test data sets.
- B. The sample means from the training data set are applied to the validation and test data sets.
- C. The sample means from the test data set are applied to the training and validation data sets.
- D. The sample means from each partition of the data are applied to their own partition.

Answer: B

Explanation:

QUESTION NO: 3

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. modelresp(event= '1') = gender region/outroc=roc;
- D. modelresp(event="1") = gender region/ out=roc;

Answer: B

Explanation:

QUESTION NO: 4

In partitioning data for model assessment, which sampling methods are acceptable? (Choose two.)

- A. Simple random sampling without replacement
- B. Simple random sampling with replacement
- C. Stratified random sampling without replacement
- D. Sequential random sampling with replacement

Answer: A,C

Explanation:

QUESTION NO: 5

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;
```

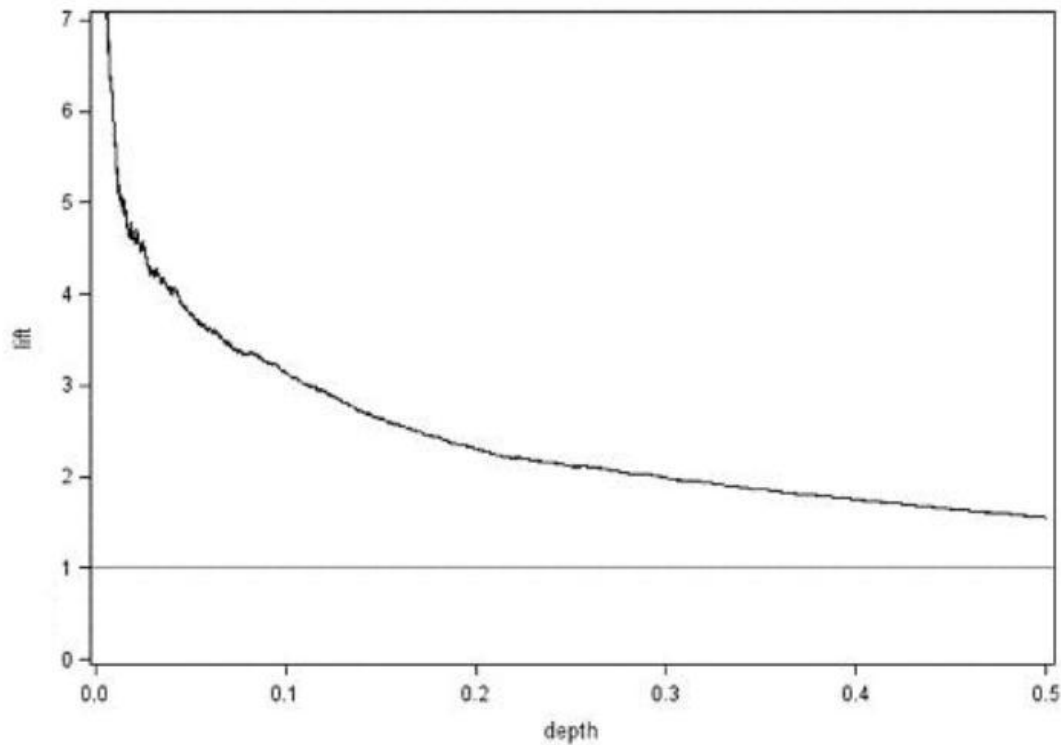
- A. Option A
B. Option B
C. Option C
D. Option D

Answer: C

Explanation:

QUESTION NO: 6

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%.

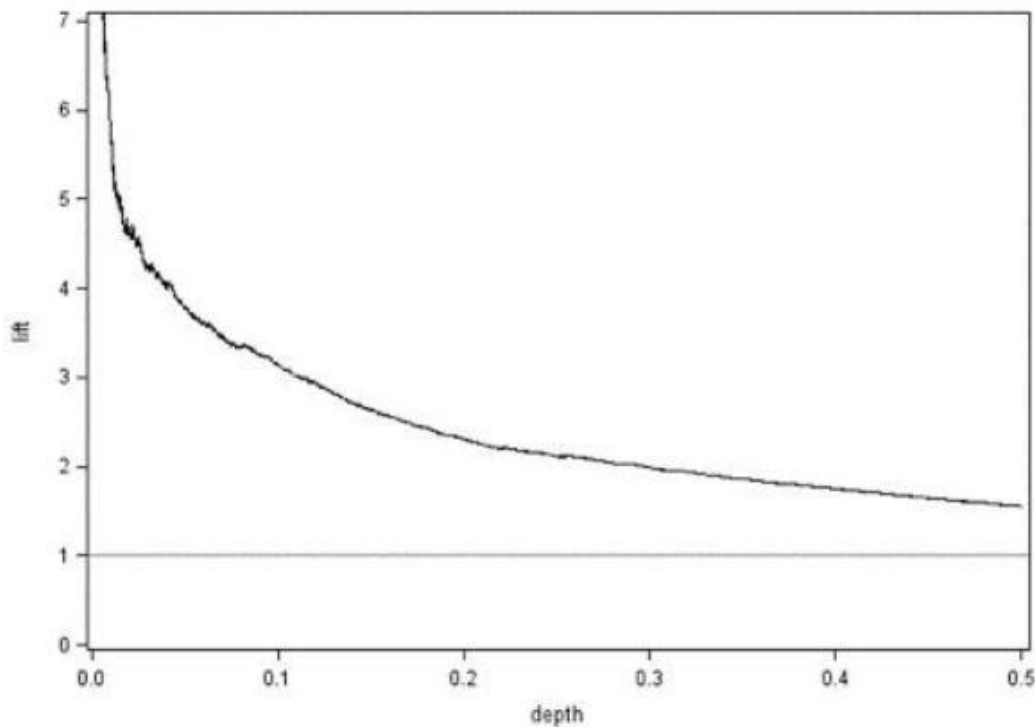
Answer: A

Explanation:

QUESTION NO: 7

Refer to the lift chart:

Refer to the lift chart:



What does the reference line at lift = 1 corresponds to?

- A. The predicted lift for the best 50% of validation data cases
- B. The predicted lift if the entire population is scored as event cases
- C. The predicted lift if none of the population are scored as event cases
- D. The predicted lift if 50% of the population are randomly scored as event cases

Answer: B

Explanation:

QUESTION NO: 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

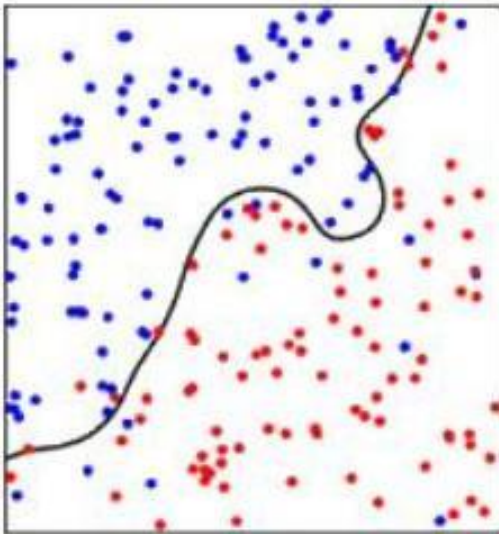
Explanation:

QUESTION NO: 9

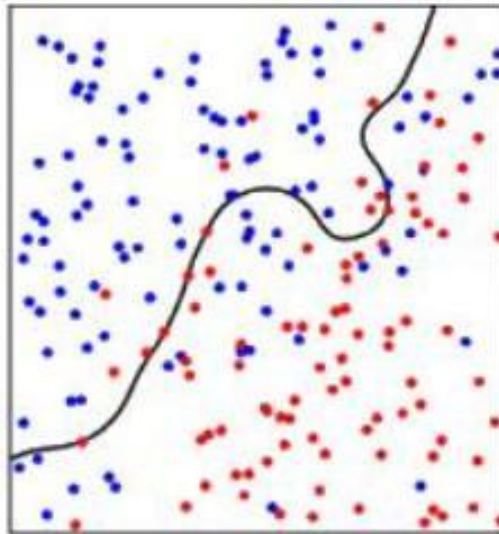
Refer to the exhibit:

Model A

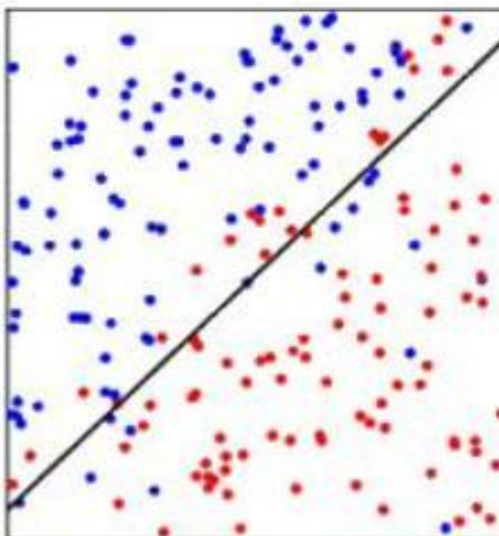
training data



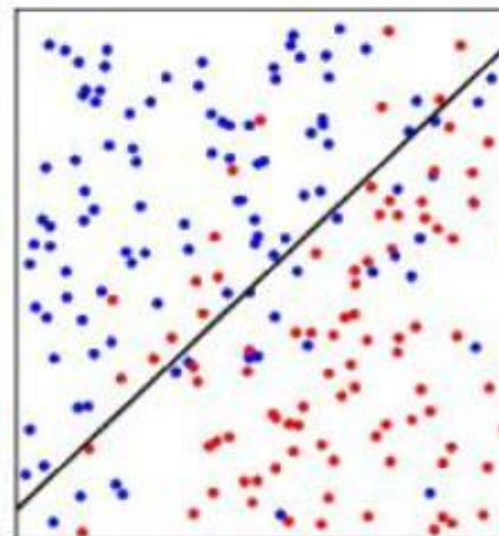
validation data

**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.

Answer: D

Explanation:

QUESTION NO: 10

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

Explanation:

QUESTION NO: 11

In order to perform honest assessment on a predictive model, what is an acceptable division between training, validation, and testing data?

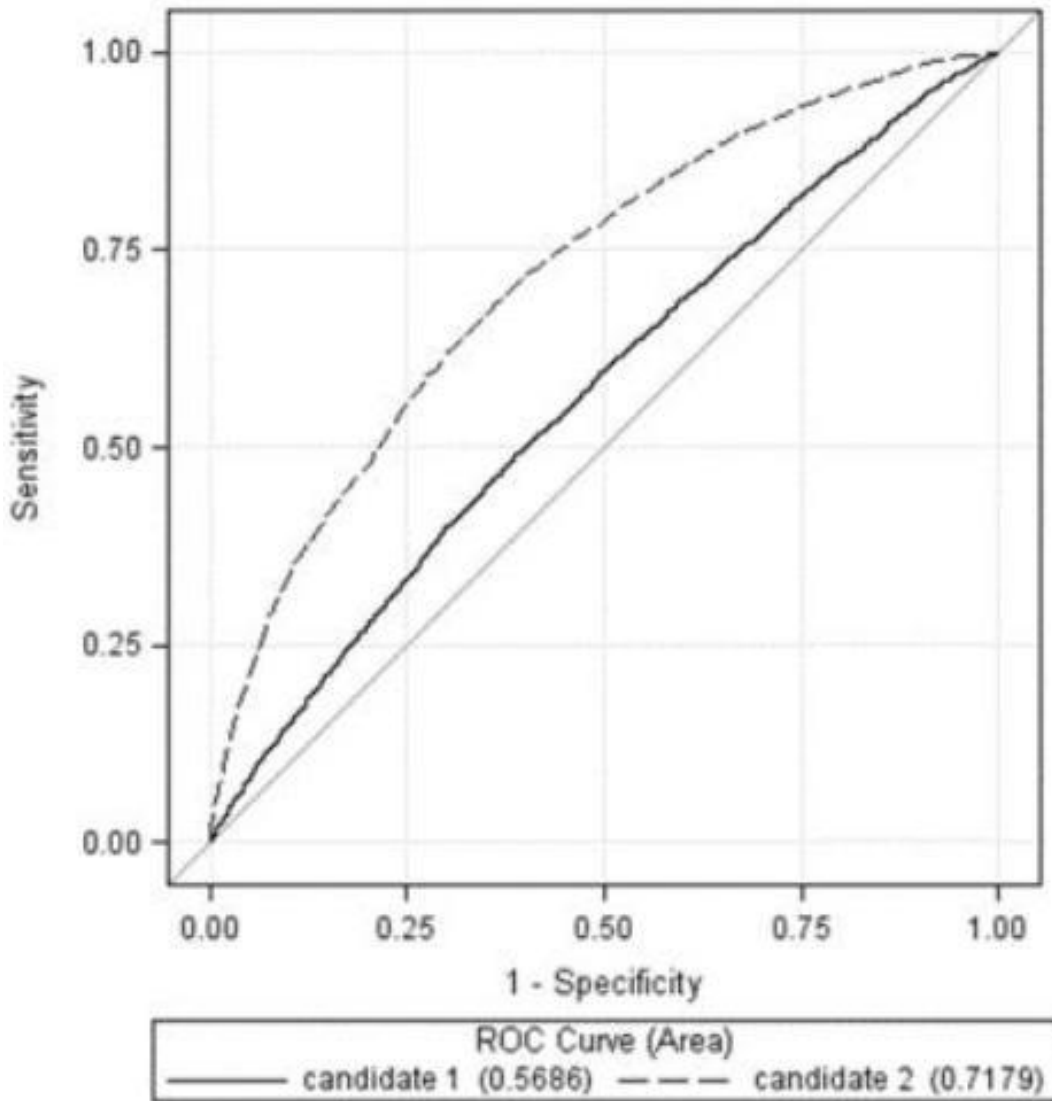
- A. Training: 50% Validation: 0% Testing: 50%
- B. Training: 100% Validation: 0% Testing: 0%
- C. Training: 0% Validation: 100% Testing: 0%
- D. Training: 50% Validation: 50% Testing: 0%

Answer: D

Explanation:

QUESTION NO: 12

Refer to the exhibit:



Based upon the comparative ROC plot for two competing models, which is the champion model and why?

- A. Candidate 1, because the area outside the curve is greater
- B. Candidate 2, because the area under the curve is greater
- C. Candidate 1, because it is closer to the diagonal reference curve
- D. Candidate 2, because it shows less over fit than Candidate 1

Answer: B

Explanation:

QUESTION NO: 13

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.

Solicit	Purchase	
	No	Yes
No	-50	0
Yes	0	450

☐ B.

Solicit	Purchase	
	No	Yes
No	0	0
Yes	-50	500

☐ C.

Solicit	Purchase	
	No	Yes
No	0	0
Yes	-50	450

☐ D.

Solicit	Purchase	
	No	Yes
No	-50	0
Yes	0	500

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

Answer: C

Explanation:

QUESTION NO: 14

A confusion matrix is created for data that were oversampled due to a rare target.

What values are not affected by this oversampling?

- A. Sensitivity and PV+
- B. Specificity and PV-
- C. PV+ and PV-
- D. Sensitivity and Specificity

Answer: D

Explanation:

QUESTION NO: 15

This question will ask you to provide missing code segments.

A logistic regression model was fit on a data set where 40% of the outcomes were events (TARGET=1) and 60% were non-events (TARGET=0). The analyst knows that the population where the model will be deployed has 5% events and 95% non-events. The analyst also knows that the company's profit margin for correctly targeted events is nine times higher than the company's loss for incorrectly targeted non-event.

Given the following SAS program:

```
proc logistic data = LOANS descending;
  model Purch = Inc Edu;
  score data = LOANS_V out = LOANS_VS priorevent = (insert X here);
run;
data LOANS_VS; set LOANS_VS;
  Solicit = P_1 > (insert Y here);
run;
```

What X and Y values should be added to the program to correctly score the data?

- A. X=40, Y=10
- B. X=.05, Y=10
- C. X=.05, Y=.40
- D. X=.10, Y=.05

Answer: B

Explanation:

QUESTION NO: 16

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

Explanation:

QUESTION NO: 17

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

Explanation:

QUESTION NO: 18

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

Explanation:

QUESTION NO: 19

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

Explanation:

QUESTION NO: 20

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?