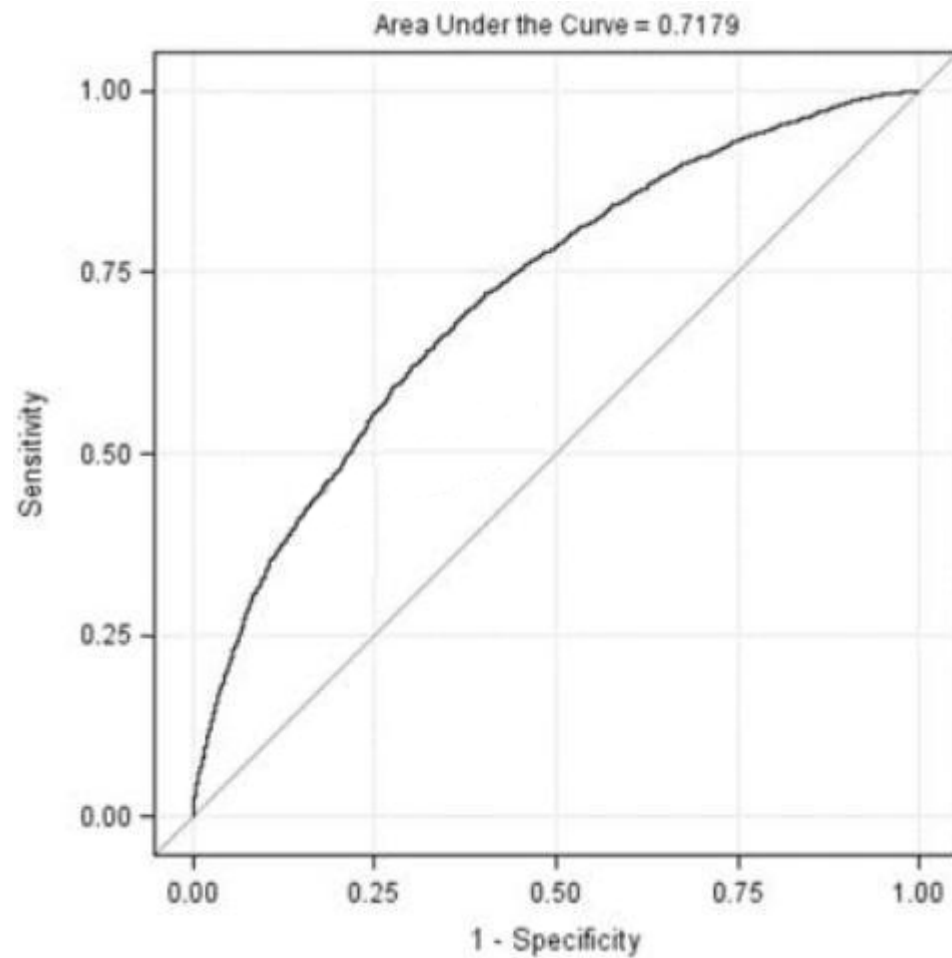**Exam Number/Code:**A00-240

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

Version: Demo

QUESTION NO: 1
Refer to the ROC curve:

Area Under the Curve = 0.7179

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

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

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. mode1resp(event= '1') = gender region/outroc=roc;
D. mode1resp(event"1") = gender region/ out=roc;

Answer: B

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

QUESTION NO: 5
Which SAS program will divide the original data set into 60% training and 40% validation data sets, stratified by county?

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

Answer: C

QUESTION NO: 6
Refer to the lift chart:
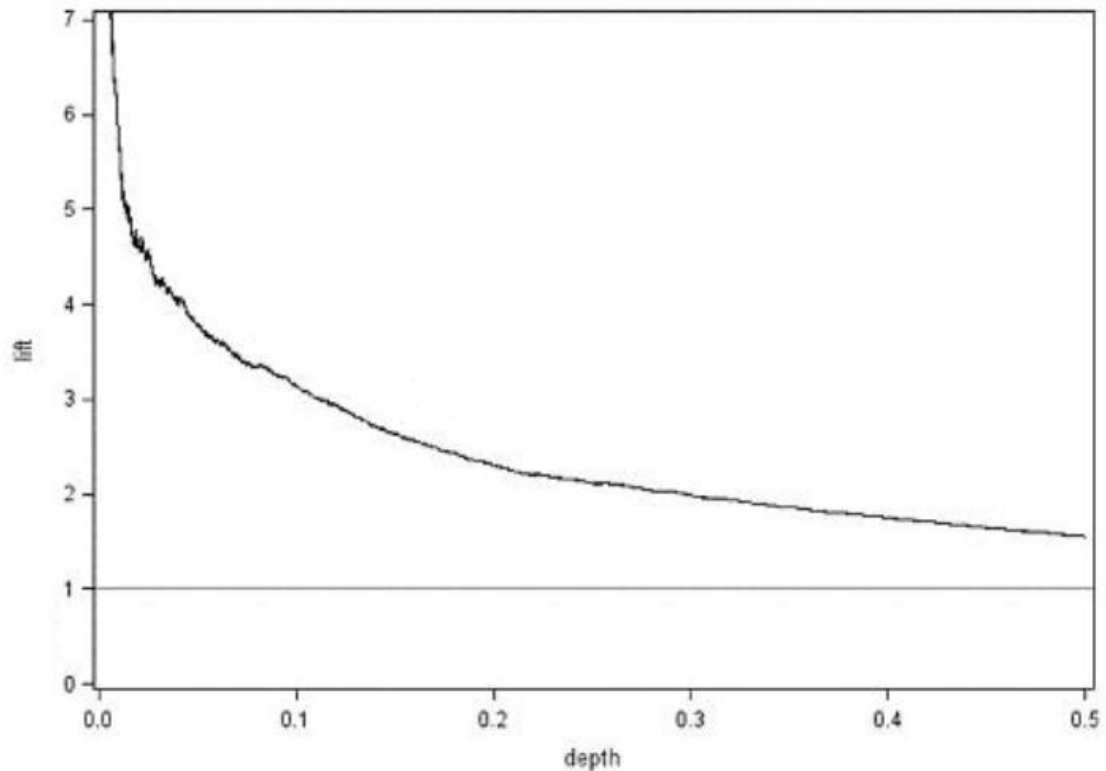
At a depth of 0.1, Lift = 3.14. What does this mean?

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 timesgreater accuracy than a random draw of 10%.
D. Selecting the observations with a response probability of atleast 10% should result in 3.14times greater accuracy than a random draw of 10%.

Answer: A

QUESTION NO: 7
Refer to the lift chart:

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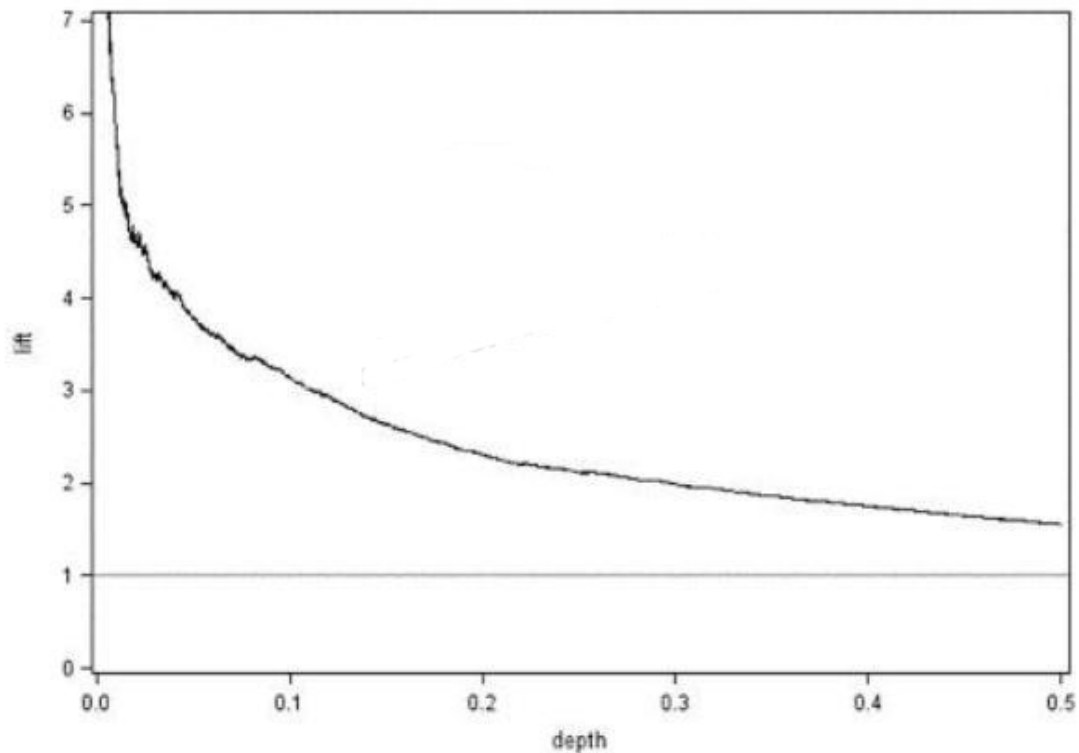
What does the reference line at lift = 1 corresponds to?

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

QUESTION NO: 8
Suppose training data are oversampled in the event group to make the number of events and nonevents 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;
```
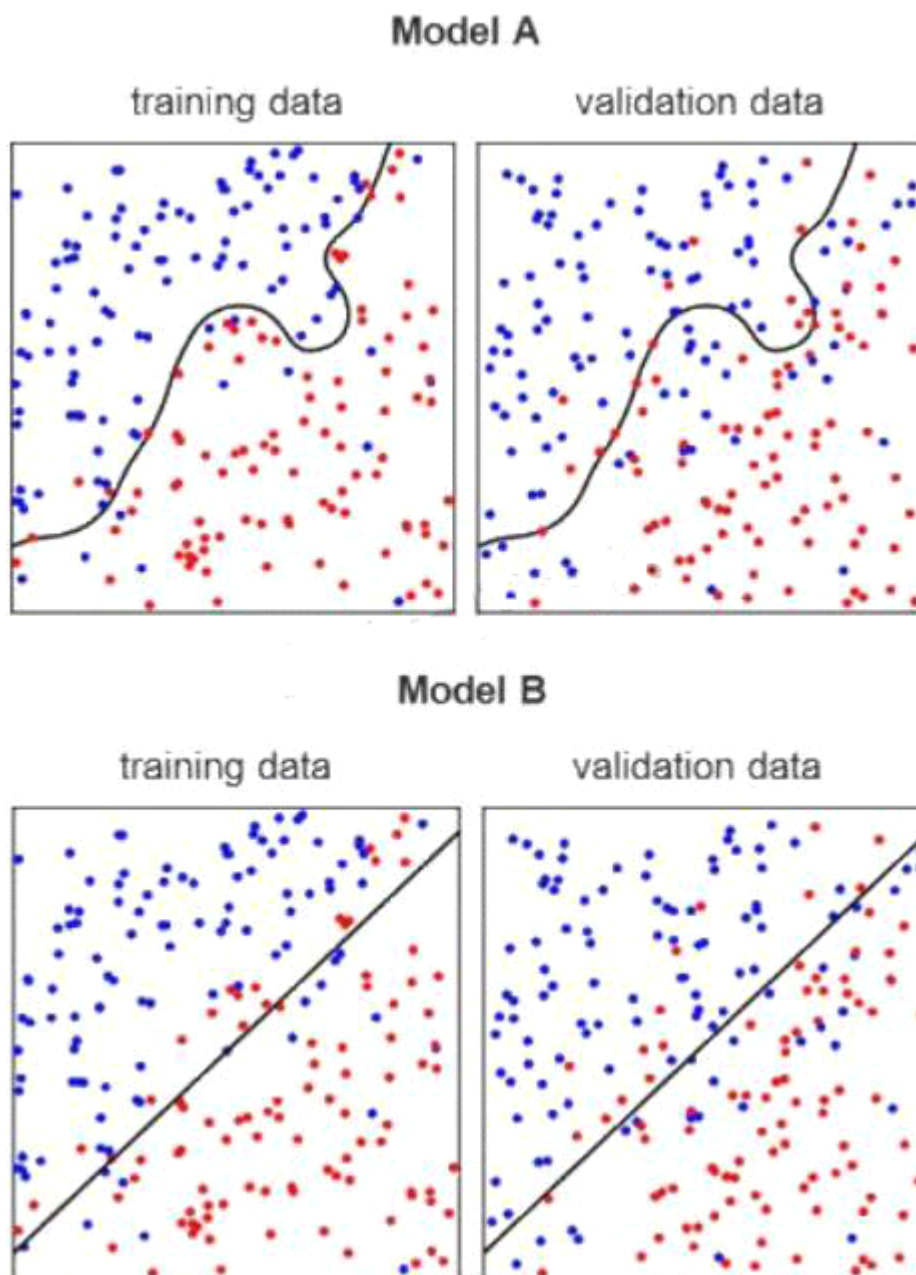
What does this program calculate?

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

Answer: B

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

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