

Sample Questions

The following sample questions are not inclusive and do not necessarily represent all of the types of questions that comprise the exams. The questions are not designed to assess an individual's readiness to take a certification exam.

SAS Statistical Business Analysis Using SAS 9: Regression and Modeling

Question 1

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.

correct_answer = "C"

Question 2

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;`

correct_answer = "B"

Question 3

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;
```
- D.

```
proc reg data=SASUSER.MLR;
  model y = c1 x1 x1*x1 c1*x1;
run;
```

correct_answer = "C"

Question 4

Refer to the REG procedure output:

| <i>Parameter Estimates</i> | | | | | | |
|----------------------------|-----------|---------------------------|-----------------------|----------------|--------------------|------------------------------|
| <i>Variable</i> | <i>DF</i> | <i>Parameter Estimate</i> | <i>Standard Error</i> | <i>t Value</i> | <i>Pr > t </i> | <i>Standardized Estimate</i> |
| <i>Intercept</i> | 1 | 618.44051 | 40.03665 | 15.45 | <.0001 | 0 |
| <i>overhead</i> | 1 | 4.99845 | 0.00157 | 3181.24 | <.0001 | 0.99993 |
| <i>scrap</i> | 1 | 2.82667 | 0.71581 | 3.95 | <.0001 | 0.00124 |
| <i>training</i> | 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

correct_answer = "B"

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.

correct_answer = "B"

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.

correct_answer = "D"

Question 7

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. 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.

correct_answer = "C & D"

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

correct_answer = "A"

Question 9

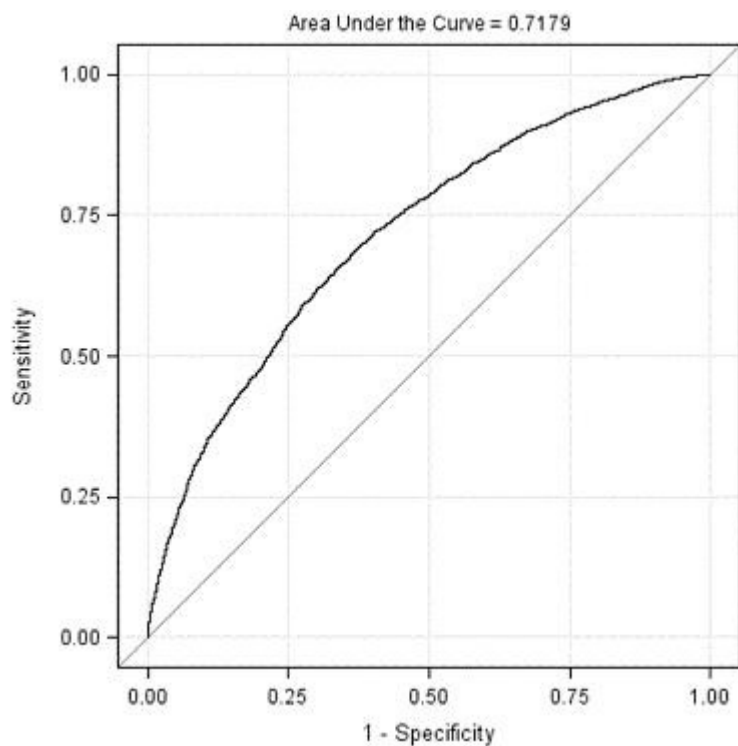
When mean imputation is performed on data after the data is partitioned for honest assessment, what is the most appropriate method for handling the mean imputation?

- 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.

correct_answer = "B"

Question 10

Refer to the exhibit:



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

correct_answer = "B"