

サンプル問題

以下のサンプル問題は包括的ではなく、必ずしも試験を構成するすべての種類の問題を表すとは限りません。問題は、個人が認定試験を受ける準備ができているかどうかを評価するためのものではありません。

SAS Statistical Business Analysis Using SAS 9:Regression and Modeling

問題1

財務アナリストは、ポートフォリオAの資産がポートフォリオBの資産よりもリスクが高い（分散が大きい）かどうかを知りたいとします。アナリストは、2つのグループのそれぞれの資産の年間収益（または変化率）を計算し、GLMプロシジャから次の出力を取得します。

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		

どの結論が出力によって支持されますか。

- A. ポートフォリオAの資産は、ポートフォリオBの資産よりもかなり危険です。
- B. ポートフォリオBの資産は、ポートフォリオAの資産よりもかなり危険です。
- C. ポートフォリオはリスクに関して大きく異なる。
- D. ポートフォリオはリスクに関して大幅に異なる。

正解 = "C"

問題2

アナリストは、地域によって大きな影響があると判断しました。アナリストは、8つの地域すべてをペアごとの比較をする必要があります、実験的なエラー率を制御したいと考えています。

どのGLMプロシジャステートメントが正しい出力を提供しますか。

- 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;

正解 = "B"

問題3

ある線形モデルは、以下の特徴を持っています。

- 従属変数 (y)
- 2次項 (x1²) を含む1つの連続値の予測変数 (x1)
- 1つのカテゴリカルな予測変数 (c1 with 3 levels)
- 1つの交互作用項 (c1 by x1)

どのSASプログラムがこのモデルに適合していますか？

- 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;`

正解 = "C"

問題4

REGプロシジャの出力を参照してください。

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

応答変数の最も重要な予測子は何ですか。

- A. intercept
B. overhead
C. scrap
D. training

正解 = "B"

問題5

ロジスティック回帰の仮定として正しい記述はどれですか。

- A. サンプルサイズは100を超えます。
- B. ロジットは予測変数の線形関数です。
- C. 予測変数は相関しません。
- D. 誤差は通常は分散されます。

正解 = "B"

問題6

LOGISTICプロシジャでSELECTION = BACKWARDを使用して変数または効果を選択すると、ビジネスアナリストのモデル選択が第3ステップで終了しました。

第1ステップと第2ステップの間に何が起こったと考えられますか。

- A. DFが増加した。
- B. AICが増加した。
- C. Pr> Chisqが増加した。
- D. - 2 Log Lが増加した。

正解 = "D"

問題7

予測モデルで、複数の変数で欠損値があるデータセットを使用します。

このモデルにはどのような問題が2つあると考えられますか。（2つを選択してください）

- A. モデルはおそらくオーバーフィットします。
- B. 入力変数間には共線性の割合が高くなります。
- C. モデル構築のプロセスにおいて、より少ない数のオブザベーションが使用されます。
- D. 入力変数に欠損値が存在する新しいケースは、余分なデータ処理なしではスコアリングできません。

正解 = "C&D"

問題8

アナリストは、各入力とターゲット変数との間の関連性の強さを推定することによって、無関係な変数をスクリーニングしようとしています。アナリストは、CORRプロシジャでSpearmanの相関とHoeffdingのD統計量を使用しています。

大きなHoeffdingやSpearmanの統計値がほぼゼロになるいくつかの入力変数は、何によって生じる可能性があるでしょうか。

- A. 変数間の非単調な関係
- B. 変数間の線形関係
- C. 変数間の単調な関係
- D. 変数間の関係が見られない

正解 = "A"

問題9

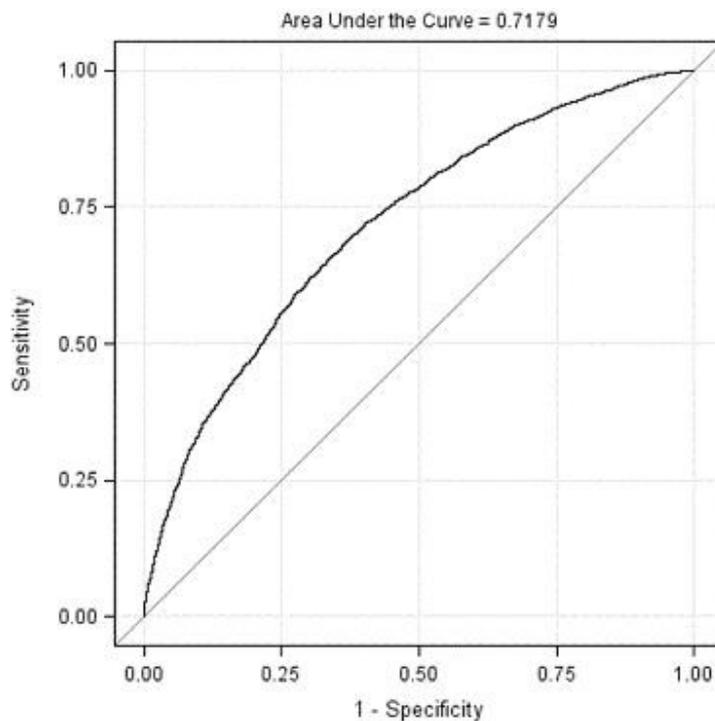
「正直な評価」のために、データを分割した後にデータの平均値による補完を行う場合、平均値による補完を処理するための最も適切な方法は何ですか。

- A. 検証データセットからの標本平均を、学習およびテストデータセットに適用します。
- B. 学習データセットからの標本平均を、検証およびテストデータセットに適用します。
- C. テストデータセットの標本平均を、学習および検証データセットに適用します。
- D. データの各分割からの標本平均を、それぞれの分割に適用します。

正解 = "B"

問題10

図を参照してください。



示されているROC曲線では、曲線下の面積は何を意味していますか。

- A. 一致の割合 + タイの割合
- B. 一致の割合 + (0.5 * タイの割合)
- C. 一致の割合 + (0.5 * 不一致の割合)
- D. 不一致の割合 + タイの割合

正解 = "B"