

Контроль набутих знань практика 5

Завдання

Для роботи використовується файл Titanic.csv

Список обов'язкових кроків для виконання:

1. Підготовка даних

- 1.1. Створіть код в SAS Studio з назвою у форматі - lastname_cw8.sas;
- 1.2. Створіть у SAS Studio макрозмінні для свого SAS ID та прізвища.
- 1.3. Створіть бібліотеку(назва бібліотеки=прізвище інтерна), в якій будете зберігати лише фіналізовані датасети (*проміжні датасети зберігайте в директорії WORK*).

2. Формування датасету для аналізу

- 2.1 Імпортуйте датасет titanic.csv
- 2.2 Створіть нумерік змінну sex на основі змінної gender з форматом “female”=1 і “male” = 2
- 2.3 використовуючи call streaminit і rand("Uniform") стерти деякі значення як показано в таблиці нижче

Сід для call streaminit	Якщо rand("Uniform") менше за	Зробити місінговим
2024	0.15	fare
2025	0.2	class
2026	0.25	sex

- 2.4 Непарні обсервації збережіть в датасет Titanic1, а парні - в Titanic2.

3. Побудова логістичних моделей

- 3.1 Створити макрос, який буде робити логістичні моделі (зберігати датасет ParameterEstimates з процедури proc logistic) з вказаним набором даних і моделлю і зберігати в вказаний датасет, за замовчуванням в ньому прописано ods exclude all, яке можна вимкнути за потреби.

Використовувати (event='1').

Збережений датасет має наступний вигляд:

	Variable	DF	Estimate	StdErr	WaldChiSq	ProbChiSq	_ESTTYPE_
1	Intercept	1	7.8002	1.1151	48.9297	<.0001	MLE
2	class	1	-1.3293	0.2586	26.4184	<.0001	MLE
3	sex	1	-2.7196	0.3437	62.6061	<.0001	MLE
4	age	1	-0.0306	0.0136	5.0955	0.0240	MLE
5	fare	1	-0.00266	0.00313	0.7198	0.3962	MLE

3.2 використати макрос щоб отримати наступні датасети, використовуючи датасет Titanic1:

Назва вихідного датасету	Модель
out1111	Survived= class sex age fare
out1110	Survived= class sex age
out1101	Survived= class sex fare
out1011	Survived= class age fare
out0111	Survived= sex age fare
out1100	Survived= class sex
out1010	Survived= class age
out0110	Survived= sex age
out1001	Survived= class fare
out0101	Survived= sex fare
out0011	Survived= age fare
out1000	Survived= class
out0100	Survived= sex
out0010	Survived= age
out0001	Survived= fare

4. Створення функції

4.1 Створити функцію (proc fcmp) на основі датасетів отриманих в попередньому розділі: функція буде мати змінні class sex age fare і на основі цих значень повертатиме ймовірність вижити особи.

Функція використовує коефіцієнти тієї моделі, для якої є немісінгові значення (тобто якщо вік місінговий, а все інше відомо, то беруться значення з моделі в out1101).

Якщо всі вхідні значення місінгові, то і результат буде місінговий.

При побудові функції коефіцієнти з моделей out: заокруглити до тисячних

Формула для побудови логістичної регресії: $\frac{1}{1+\exp(-(\theta_0+\theta_1x_1+\theta_2x_2+\dots))}$, де θ - коефіцієнти, x - змінні. θ_0 - коефіцієнт зсуву (Intercept).

4.2 Створити датасет `P_Titanic` використавши датасет `Titanic2`, в який додати змінну `p_survived` створеною функцією.

5. Аналіз результату

5.1 Проаналізуйте змінні `survived` і `p_survived` в датасеті `Titanic2`, зробіть висновки.