

Контроль набутих знань

Практика 1

Завдання

Вхідний набір даних Titanic містить інформацію про пасажирів, потрібно написати програму, що друкує звіт із використанням **PROC PRINT**.

Для роботи використовується файл Titanic.csv.

Приклад звіту є в Додатку 2, до цього документу.

Список обов'язкових кроків для виконання:

1. Ознайомитись з інформацією про набір даних Titanic (Додаток 1) ;
2. Створити код в SAS Studio і назвіть lastname_cw1;
3. Створити бібліотеку(назва бібліотеки=прізвище інтерна), в якій будете зберігати всі проміжні датасети.
4. Імпортувати файл Titanic.csv в свою бібліотеку:
 - Базовий варіант виконання: використання PROC IMPORT (1 бал);
 - Складніший варіант виконання, імпорт через Data Step (4 бали);

Доступно 2 варіанти виконання, використовуєте лише один!

5. Використовуючи **PROC CONTENTS** перегляньте інформацію про датасет Titanic, і вивгрузіть метадані в датасет **meta**. Датасет meta повинен бути відсортований за порядком змінних у датасеті Titanic, і містити лише наступні змінні: Variable/Type/Len (див.зразок нижче);

| Variable | Type | Len |
|----------|------|-----|
| Name | Char | 51 |
| Age | Num | 8 |
| Gender | Char | 6 |
| Class | Num | 8 |
| Fare | Num | 8 |
| Survived | Num | 8 |

6. Для побудови звіту необхідно:

а) вибрати вибірку суб'єктів за наступною умовою:

- пасажир їхав 1 класом і залишився живим;
- або вартість квитка більша рівна 50.

б) після чого виключити зі датасету змінні Survived і Fare;

в) додати змінну AgeGroup (Child: до 12, Teen: 13–18, Adult: 19+, Unknown – missing values).

г) створити змінну Sex, яка буде мати значення ("F"= "Female", "M"= "Male") і довжину 30.

д) Створіть нову змінну Title, яка буде містити лише титул пасажера, вилучений з імені (наприклад: Mr, Mrs, Miss, Master і т.д.).

е) відсортувати датасет за наступним ключем: клас за спаданням, вік і ім'я за зростанням

7. Побудуйте звіт використовуючи сформований фінальний датасет (**REPORT**) за допомогою PROC PRINT (виведіть лише перші 30 спостережень)

Загалом, можливі деякі моменти, що не досить формалізовані в постановці завдання, в цьому випадку раджу проявити ініціативу та креативність, яку опишіть у вигляді коментаря в початку вашої програми.

Додаток 1. Теоретична інформація.

В якості прикладу розглянемо, досить популярний серед статистиків набір даних під назвою **Titanic**.

Опис даних Titanic.

10 квітня 1912 року "Titanic" вирушив в своє перше плавання через Атлантичний океан, на борту було 2223 пасажирів. 14 квітня він врізався в айсберг та затонув. Загибло 1517 людей. В наявності була інформація не про всіх пасажирів.

Опис полів набору даних Titanic:

Survived – якщо пасажир вижив то Survived=1, якщо загинув то Survived =0.

Age – вік пасажера в роках.

Gender – стать пасажера (male - чоловіча, female - жіноча).

Class - яким класом їхав пасажир (1, 2, 3).

Fare – вартість квитка. Ця змінна зазвичай не використовується, тому що відображає вартість всіх квитків в каюті, а не індивідуального місця.

Цей набір даних, є публічним для використання у вільному доступі.

Додаток 2. Приклад звіту.

Прізвище ім'я інтерна

Дата побудови звіту

Заголовок звіту (по центру, наприклад, Звіт про пасажирів Титаніка)

| Name | Age | Gender | Class | AgeGroup | Sex | Title |
|------|-----|--------|-------|----------|-----|-------|
| | | | | | | |
| | | | | | | |

Примітка: у виведеному уривку звіту присутні __ суб'єктів з _ класу, в яких відсутня інформація про вік. Виведено _ суб'єктів, серед яких _ дітей, _ підлітків та _ дорослих.