

Контроль набутих знань

Практика 2

Завдання

Для роботи використовується файли: Titanic.csv, test.csv, train.csv і titanic3.xls.

Список обов'язкових кроків для виконання:

1. Ознайомитись з наборами даних.
2. Створити код в SAS Studio і назвати lastname_cw2;
3. Створити макро-змінні:
 - Зі своїм SAS ID (приклад-u57972979);
 - Зі своїм прізвищем.Далі ці макрозмінні необхідно використовувати в коді.
4. Створити бібліотеку(назва бібліотеки=прізвище інтерна), в якій будете зберігати всі проміжні датасети. Відповідно тут потрібно вже використати попередньо створену макрозмінну.
5. Імпортуєте всі необхідні файли в свою бібліотеку. (додатково - створити макропрограму, яка автоматично імпортує всі CSV-файли з певної директорії)
6. Файли train.csv і test.csv це дві частини одного файлу.
7. До датасету Titanic, з датасетів test і train потягніть змінні PassengerId, sibsp та parch.
8. До датасету Titanic, з датасету titanic3 потягніть змінні embarked boat body home_dest.
9. Значення змінної embarked приведіть у відповідність з Додатком 1 (C = Cherbourg; Q = Queenstown; S = Southampton).
10. Створіть нову змінну FamilySize, яка розраховується як сума sibsp та parch + 1 (сам пасажир). Категоризуйте її на:
 - Small (1-2)
 - Medium (3-4)
 - Large (5 і більше).
11. Створіть змінну Category за наступним алгоритмом. Далі збережіть даний датасет як **REPORT**, відсортуйте його за ключом NAME, і виключіть змінні AGE і Survival.

Category	Age	Survival
Child - Survived	0-12	Yes
Child - Deceased	0-12	No
Teen- Survived	13-18	Yes
Teen- Deceased	13-18	No
Adult- Survived	19+	Yes
Adult- Deceased	19+	No

12. Назва змінних і лейбли у фінальному датасеті (**REPORT**) мають відповідати Додатку 1. Якщо в додатку ці змінні(або аналоги) не представленні, то назвіть за власним бажанням.
13. Використовуючи **PROC CONTENTS** виграйте метадані з датасету **REPORT** в датасет **META**. Датасет meta повинен містити лише наступні змінні: Variable/Label/Type/Length а також створіть змінну (Source) в якій вкажіть походження змінної.(приклад – Input data set Titanic/Test/Train/Titanic3, Created etc).
14. Порахуйте на Дата кроці осіб в кожній категорії. Створіть датасет **FREQ** який містить 2 змінні (назва категорія і кількість осіб).
15. Створіть датасет **MATRIX** який матиме наступний вигляд:

NAME	Child_Survived	Child_Deceased	Teen_Survived	Teen_Deceased	Adult_Survived	Adult_Deceased
XXXX	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No	Yes/No

16. За допомогою PROC PRINT створіть репорти з датасетів **META/FREQ/ MATRIX/REPORT** додавши тайтли і футноут.
Додаткові завдання:
17. Створіть код, що перевіряє наявність певних змінних у датасеті та виводить інформацію в журнал SAS, якщо змінної немає.
18. Використовуючи PROC SQL, створіть новий датасет, який міститиме лише пасажирів першого класу із квитком дорожче середнього.
19. Побудуй розподіл пасажирів по FamilySize у вигляді стовпчикової діаграми

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.

VARIABLE DESCRIPTIONS

Pclass	Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
Survival	Survival (0 = No; 1 = Yes)
name	Name
Sex	Sex
age	Age
sibsp	Number of Siblings/Spouses Aboard
parch	Number of Parents/Children Aboard
ticket	Ticket Number
fare	Passenger Fare (British pound)
cabin	Cabin
embarked	Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
boat	Lifeboat
body	Body Identification Number
home.dest	Home/Destination

SPECIAL NOTES

Pclass is a proxy for socio-economic status (SES)

1st ~ Upper; 2nd ~ Middle; 3rd ~ Lower

Age is in Years; Fractional if Age less than One (1)

If the Age is estimated, it is in the form xx.5

Fare is in Pre-1970 British Pounds ()

Conversion Factors: 1 = 12s = 240d and 1s = 20d

With respect to the family relation variables (i.e. sibsp and parch) some relations were ignored. The following are the definitions used for sibsp and parch.

Sibling: Brother, Sister, Stepbrother, or Stepsister of Passenger Aboard Titanic

Spouse: Husband or Wife of Passenger Aboard Titanic (Mistresses and Fiances Ignored)

Parent: Mother or Father of Passenger Aboard Titanic

Child: Son, Daughter, Stepson, or Stepdaughter of Passenger Aboard Titanic

Other family relatives excluded from this study include cousins, nephews/nieces, aunts/uncles, and in-laws. Some children travelled only with a nanny, therefore parch=0 for them. As well, some travelled with very close friends or neighbors in a village, however, the definitions do not support such relations