**The University of Edinburgh**

**Business School**

**Credit Risk Management**

**Computer Lab I. Initial data exploration and coarse-classification**

Download the data in “*Development\_sample.csv*” from Learn. The file contains a simulated sample from the accepted population of credit card applicants. Save the dataset into your desired place and remember to save all the output at the end of each computer session.

There are 12 variables:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| ID | Account number, unique identifier |
| FLAG BINARY | Performance flag: 0- Bad, 3 or more missed payments; 1 - Good, less than 3 missed payments |
| PHONE | 0 - No telephone number given; 1 - Home number given; 2 - Other number given |
| MSTAT | Marital status: 1- Married; 2- Single; 3- Divorced; 4- Widowed; 5 - Co-habiting |
| RSTAT | Residential status: 1- Home Owner; 2 - Renting House; 3- Renting Flat; 4 - Social Housing; 5- With Parents; 6 - Unknown |
| KIDS | No of children |
| CARDI | 0 - No Card Insurance Taken; 1- Card Insurance Taken |
| CARDH2 | 0 - No 2nd Card Holder; 1 - 2nd Card Holder |
| AGE | Applicant's age in years |
| TAD | Time at address in months |
| BUREAU | Number of negative records at credit bureau |
| SEARCH | Number of credit searches in the last 3 months |

1. **Explore the content of the data.**
2. **Explore categorical characteristics (e.g., PHONE).**

*Exercise*: Obtain bar chart and frequency table for other nominal variables (MSTAT, RSTAT, KIDS, CARDI, CARDH2, BUREAU and SEARCH).

1. **Association with Performance Flag (e.g., PHONE).**

a). Obtain a cross-table between variable PHONE and the performance flag FLAGBINARY.

b). Obtain the measures of association between the variables. Interpret the Chi-square statistic.

1. **Creating new coarse-classified variables (e.g., PHONE).**

a). If we decide to merge categories ‘no number given” and ‘other number given’, we need to create a new variable, the coarse-classified variable ‘Phone1’.

b). Repeat cross-tabulation as described in section 3 for ‘Phone1’ and performance flag FLAGBINARY.

c). Alternatively, one may decide to merge categories ‘home number given’ and ‘other number given’. Please, create a new variable called ‘Phone2’ by merging these two categories and get the association measures for this new variable and the performance flag.

What merging is better and why?

1. **Weights of Evidence.**

Use cross-tabulation information to calculate WoE for ‘Phone1’. Using the method described in 4a) create a new variable (e.g., PhoneWoe)

1. **Further Practice.**

Now repeat the same procedure, from section 2 to section 5, for marital status ‘MSTAT’. Explore the variable. Get a table of frequencies for the variable and a bar chart. Obtain the cross-tab with the performance flag. Create grouped variables and recode them to dummy and WoE variables.

**Homework**

Repeat coarse-classing for all categorical variables in the dataset. Transform coarse-classed characteristics into dummy and WOE variables. Remember that you need k-1 dummies for k coarse classes! Remember to save your dataset with all new variables.