**The University of Edinburgh**

**Business School**

**Credit Risk Management**

**Computer Lab 2. Coarse-classification of numeric characteristics**

The objective of Labs 1 & 2 is to practise the stage ‘Characteristics analysis’ (see Session 2). This stage is a compulsory stage in scorecard development. This stage is necessary because before regression analysis, one needs to convert the original variables into the form which would make further analysis possible. For more in-depth discussion of reasons behind coarse-classification (variable transformation) see Session 2 slides and associated reading. It is also recommended to refresh your knowledge of Session 1 (Naïve Bayes example, where the concept of WoE was first introduced).

Lab 1 has provided instructions for coarse-classifying categorical variables. Lab 2 extends the procedure to numeric variables. After Computer lab 1 you should have “*Devsample\_coarse\_3.csv*” with the following new variables:

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Phone1 | 1 – Home number given  0 – Otherwise |
| PhoneWoe | 0.08 – Home number given  -0.79 – Otherwise |
| MSgrouped | 1 – Married or Widowed  2 – Single  3 – Divorced or Co-habiting |
| Married1 | 1 – Married or Widowed  0 – Otherwise |
| Married2 | 1 – Single  0 – Otherwise |
| MSWOE | 0.47 – Married or Widowed  -0.50 – Single  -0.06 – Divorced or Co-habiting |

If you have used different variable names, it is fine.

1. **Exploring numeric characteristics (e.g., AGE).**

a). Examine descriptive statistics and histograms of AGE.

b). Split AGE into initial fine classes to prepare them for coarse-classification.

c). Obtain a univariate frequency table of the new variable. You can examine if the bins have approximately the same number of observations. Note that because of the discrete nature of “Age” the frequency differs slightly across the bins.

1. **Association with Performance Flag and coarse-classification.**

a). Cross-tabulation. Now we want to examine the cross-tabulation frequency table for “agebinned” and the “flag”.

b). Creating new coarse-classified variables and converting them into dummies.

**Further practice**

Coding Age with Weights of Evidence (WOE). Use the same method and formula as in Computer Lab 1 to create a new variable AgeWoe from AgeGrouped. Note than you have to calculate WoE for each category yourselves using the numbers from a cross-tab of AgeGrouped\* Flag.

Repeat the same procedure for ‘Time at Address’.