

**Exercise Sheet 3, Business Analytics, SoSe 2011, 30.05.2011**  
**Dr. Tomáš Horváth, Osman Akcatepe**

1. Normalize the attribute X by
  - (a) decimal scaling if it ranges between 0.88 and 999.99, when we set j is equal to 3,
  - (b) again decimal scaling if the range of attribute X is -500 to 45 with j=3.
  - (c) min-max normalization to range of the interval [0,1] if its value is 40 between the interval [20,100] which is needed to be transformed.
2. Suppose the data  $D = \{13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70\}$  for analysis includes the attribute Age. Use binning with mean to smooth the above data, using a bin depth of 3. Illustrate your steps.
3. Using the data D for Age in Question 2, answer the following:
  - (a) Use min-max normalization to transform the value 35 into the range [0.0, 1.0].
  - (b) Use z-score normalization to transform the value 35, standard deviation is 12.94.
  - (c) Use normalization by decimal scaling to transform the value 35 for age.
  - (d) Comment comparing the normalization methods above as to why on which you would prefer to use for the given data.
4. Using the data for Age in Question 2, give an example of stratified sampling using proportional allocation. Use samples of size 13 and the groups as
  - (a) Age < 22 and
  - (b) Age ≥ 22.

5. Assume that  $N = 10$ ,  $\phi(i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ ,  $N\left(\frac{2N}{5}, \frac{N}{10}\right)$ ,  $f(n)=n^2$ . Find the minimal costs and compute the schedules corresponding to minimal costs.

Bonus: Write a basic program in order to compute the schedule with  $N= 100, 1000, 10K, 100K$  with above given distribution and function. How does result seem?

*Good Luck !*