

### Assignment 3

**Due:** 9/29

**Note:** Show all intermediate calculations/results. You can do manual calculations or use any software (e.g., Weka, Excel, JMP, R) to answer the questions, unless otherwise noted. In any case, you need to attach the relevant file(s) or screenshot(s) that shows how you obtained your answers.

**Problem 1 (20 points)** Consider the following dataset (sorted in non-decreasing order):  
<13, 38, 41, 44, 45, 81, 83, 88, 90, 95, 103, 125, 129, 137, 152>

- (1) Perform the equal width binning on the above data with 3 bins using the method we discussed in the class. For each bin, show the bin interval, data values in the bin, and smoothed values using bin means, bin medians, and bin boundaries.
- (2) Repeat the same with equal depth binning with 3 bins.
- (3) If you transform the dataset into the interval of [0, 10] using Min-max normalization, what is the new value of 125?
- (4) If you transform the dataset using z-score normalization using the standard deviation, what is the new value of 125?
- (5) If you transform the dataset using z-score normalization using the mean absolute deviation, what is the new value of 125?

Note: For Problem 1-(4) and Problem 1-(5), you need to show the mean, standard deviation, mean absolute deviation, and the new, transformed value as well as all calculation steps.

**Problem 2 (10 points)** This problem is a practice of calculating Pearson's correlation coefficients between predictor variables and the predicted variable in a dataset. The dataset to be used for this problem is *concrete-strength-reduced.csv*, which was downloaded from the UCI Machine Learning Repository ([UCI Machine Learning Repository: Concrete Compressive Strength Data Set](#)) and modified for this assignment.

Calculate the following correlations:

Correlation between *Cement* and *Concrete\_compressive\_strength*  
Correlation between *Blast\_Furnace\_Slag* and *Concrete\_compressive\_strength*  
Correlation between *Fly\_Ash* and *Concrete\_compressive\_strength*  
Correlation between *Water* and *Concrete\_compressive\_strength*

Which attribute has the strongest correlation with *Concrete\_compressive\_strength*?

**Problem 3 (10 points)** This problem is a practice of determining correlation between two nominal attributes using the chi-square test, which we discussed in the class. Consider the *hw3-p3.csv* dataset, which was downloaded from

- (1) Determine whether there is a correlation between attribute *gender* and attribute *readmitted*.
- (2) Determine whether there is a correlation between attribute *insulin* and attribute *readmitted*.

You can use any tool when creating contingency tables. After that, you must do all calculations, including the calculation of expected values, yourself. Use 5% significance level.

**Submission:**

Name your file *LastName\_FirstName\_HW3.doc* (or *LastName\_FirstName\_HW3.pdf*). If you have multiple files, then combine all files into a single archive file. Name the archive file as *LastName\_FirstName\_HW3.EXT*. Here, “EXT” is an appropriate archive file extension (e.g., zip or rar). Upload this archive file to Blackboard.

.