**Project Proposal: Iris Flower Data Analysis**

Yue Zhang, Jing-Yi Wu

Nancy Boudreau

Dataset URL: <https://archive.ics.uci.edu/ml/datasets/Iris>

Data description:

We want to investigate the whether we can distinguish 3 species of Iris (Iris Setosa, Iris Versicolour and Iris Virginica). On each level of Iris flower, we have 50 instances. Four variables were measured from each sample: the length of sepals, the length of petals, the width of sepals and the width of petals, in centimeters. In summary, we have 150 observations with 4 variables and 3 groups.

Questions:

1. Does the data follow multivariate normal distribution? Does the data follow univariate normal distribution for 4 variables?

2. Is there any significant difference among 3 groups? If so, where are these differences from?

3.