

Department of Statistics, School of Mathematical Sciences
Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon
M.Sc. (Statistics) I Practical Sheet: MST 506
Topic 4 Exploring Data

- 1) Suppose the letter grades of an English essay in a small class are A, D, C, D, C, C, C, C, F, and B. Create a both frequency table showing the counts and a relative frequency table showing the proportions of the various grades.
- 2) Use `quine` data frame in the `MASS` package to prepare frequency table of `Age` variable.
- 3) Construct barplots, dot charts and pie charts for the variables `Grades` used in example 1) and `Age` in the `quine` data frame from the `MASS` package using both frequencies and relative frequencies.
- 4) Use the data frame `Baberuth` from `PASWR2` package to construct a stem-and-leaf plot, strip chart and histogram for the number of home runs (`hr`) Babe Ruth hit while he played for the New York Yankees.
- 5) Construct a density histogram of the waiting time until the next eruption using the data frame `geyser` available in the `MASS` package. Superimpose a Gaussian kernel density estimate over the density histogram. In the same graph, show the kernel density estimate without showing the density histogram.
- 6) Use the data frame `Cars93` in the `MASS` package to create a boxplot of the variable `Min.Price`.
- 7) Access the data `cats` from the library `MASS` and plot sexwise box-plots for the variable `Hwt` (heart weight).
- 8) Use the data frame `Animals` from the `MASS` package to investigate whether the brain weights of animals are related to their body weights. In other words, is a bigger brain required to govern a bigger body?
- 9) Consider `Auto` data available in `ISLR2` library in R. Make sure that the missing values have been removed from the data.
 - a) Which of the predictors are quantitative, and which are qualitative?
 - b) What is the range of each quantitative predictor?
 - c) What is the mean and standard deviation of each quantitative predictor?
 - d) Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?
 - e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.
 - f) Suppose that we wish to predict gas mileage (`mpg`) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting `mpg`? Justify your answer.