Notes

Overview of linear regression

1. Data set $(x_i, y_i), i = 1, \ldots, n.$ It is desired to find out the relation between the variables X and Y. Refer to Fig 1.8 on p 16 for the strategy for regression analysis.

    (a) Preliminary step: What is the problem? How the result will be use (impact of the study)? How the data was obtained (say, nature of design)? Examination of the data: summary (descriptive) statistics, graphs.

    (b) Modeling Stage
        i. Suitable data structure for GLM (generalized linear model).
        ii. $Y = X\beta + \epsilon$. The matrix form, assumption
        iii. Suitable model, model checking and validation: diagnosis and remedial measures. Pre-data vs. Post-data inferences.

    (c) Inferences
        i. Parameter of interest: Means, Contrast of means, variances and their ratios.
        ii. Point Estimation: LSE, MLE and other methods. Their implementation and justification.
        iii. Confidence Intervals and other alternatives
        iv. Hypotheses testing: NP lemma, likelihood ratio test and justifications.
        v. Caveats of these inferences and alternatives: Parameter-centered inference, statistician-centered viewpoint
        vi. Simultaneous vs. Individual inference: Significant results can be found in most data sets.

    (d) **Thinking in the real world**: Statistical significance vs. practical significance. What is the ultimate decision and the criterion of evaluation of the decision? The limitation and scope of the results. What can go wrong?

2. Graphs and Stat figures: Ways of seeing
    (a) Statistical graphs which are commonly employed to help statisticians "see" the data: scatter plot, histogram, normal probability plot, stem-and-leaf plot, box plot.
    (b) Summary statistics: means, median, 5 points summary, quantiles, standard deviation. correlation, etc.

3. Computation utilities: what it can and what it can not help.

4. Theoretical tools: Linear Algebra (particularly matrix algebra), Calculus, Probability theory and, of course, statistical theory.