A list of class presentation topics  
Stat201

40% of your grade will be determined by the presentations you make in class during the exam week. Some possible topics are listed below, though you may like to choose your own topics outside this list. The presentations are to be done in groups. Each group should have 4-5 students. Each group will be given 45 min to present a topic. The students in the same group have to split their topic suitably into subtopics among themselves. The order in which you want to speak is up to you, but each of you should talk for no more than 10 and no less than 5 minutes.

The grading based on a class presentation cannot help being highly subjective. However, I shall be quite lenient. The main point that I shall look for in the presentations is how clearly you are presenting the basic notions, and not how many math formulas you are writing. There will be a brief question session at the end of each talk.

Most of the topics are vast, but you are to present only the basic ideas. Presenting examples is highly encouraged. I shall be happy to discuss ideas related to planning for the class presentations.

1. Logistic regression: Logistic regression is a nonlinear regression method that is designed for cases where the response variable Y takes only two possible values. The book Applied logistic regression by D. W. Hosmer and S. Lemeshow provides a nice introduction to the subject. The idea behind logistic regression is somewhat similar to that used in probit analysis.

2. Distance sampling: This is a specialized topic. It deals with a method to estimate abundance of biological populations using a special sampling scheme. It is a pretty intuitive method, and the level of math required is not much.

   The book Distance sampling: estimating abundance of biological populations by S. T. Buckland is a good reference.

   While presenting this topic you should explain the basic principle involved. Then you should detail the procedure. Give one example (you may borrow one from the book). Finally, try to suggest improvements.

3. Subset selection in regression: One method to avoid the curse of dimensionality is to use only some of the predictor variables (ie, X’s) in the model. There are different strategies to choose this “best” subset of predictors, eg, step-up and step-down regression. This presentation should give an overview and some examples of these various methods. Some useful reference are:
   - Applied Linear Regression by S. Weisberg
   - Applied multivariate data analysis (Volume I) by J. D. Jobson

4. Non-parametric tests: All our discussions in class centered around models like $X_1, \ldots, X_n$ iid with some distribution which is completely known except for some unknown parameters. Thus, the formulas of the density etc were always assumed known. In non-parametric methods, we allow the densities to have unknown formulas. There are tons of non-parametric methods. In your presentation you should pick up a handful of them (I can suggest choices), and then explain them to the class. For each method presented you should explain the intuitive idea and the exact procedure. You should give examples. Throughout our textbook, there are many examples of non-parametric procedures. You are encouraged to use these as your examples.

5. EM algorithm: Not infrequently in practice one finds that some required data is missing. Of course, it is possible to do maximum likelihood estimation using only the available data, but usually that is difficult. One modern approach is to use the Expectation-Maximization algorithm, or EM algorithm, for short. This presentation will require some math similar to what we need for finding mle’s. You should explain the idea behind the algorithm, the algorithm itself, and do one example in detail. The book The EM algorithm and extensions by G. J. McLachlan and T. Krishnan is a nice reference.
6. **Design and analysis of experiments**: Experiments constitute one main source of statistical data. In a typical statistical experiment some treatment is applied to experimental units (e.g., some drug is applied to rats) and based on the result one tries to infer the relation between the treatment and its effect. Thus designing and analysing experiments is an essential part of statistics. In this presentation you should introduce fundamental concepts like confounding effects, design matrix, randomized blocks etc. You should also present some of the statistical procedures used (e.g., what tests of hypotheses are important in this context, and how to do them). It is possible to do this presentation with a great deal of math, as well as at an intuitive level without going into much mathematical rigor. Again, our textbook by Rice is a good starting point. The book *A first course in design and analysis of experiments* by G. W. Dehler may prove helpful too.

7. **Bayesian statistics**: The type of statistics we have been doing in class is from the so-called frequentist approach, where parameters are considered as unknown fixed numbers. In Bayesian statistics one considers the parameters as random variables. It is a vast topic, and the presentation need to cover only the rudiments. Our textbook itself is a reasonably good source this material. You may also read pages 39–58 from the book *Bayesian Analysis for Linear Models* by L. D. Broemling.

8. **Statistical Quality Control**: One of the major uses of statistics in industries is statistical quality control. The basic idea is as follows. A machine is producing items one after another. When we started the machine initially it was working fine. However, we can never be sure if something has gone wrong after that (e.g., some bolt becoming loose), resulting in faulty production. To monitor the performance of the machine in the factory a few items are selected from the production line as a random sample and some statistical test is carried out to test

\[ H_0: \text{The machine is functioning properly} \]

against the alternative hypothesis that it is *not* functioning according to specifications. This is called Statistical Quality Control (SQC). The tests are often represented graphically in the form of control charts. The presentation should introduce the notion of SQC, and give some details about various control charts used. The book *Introduction statistical quality control* by D. C. Montgomery is a good reference.