INSTRUCTIONAL OBJECTIVES &
STUDY GUIDE FOR PROB & STAT FINAL EXAM
FALL 2006

I believe that you should know exactly what it is I expect you to know, and there should be no surprises in evaluating your performance. The following instructional objectives lists what I expect you to get out of this class and hence should give you a clear view of what you should know to ace the upcoming final, as well as to give you some house keeping rules regarding the exam.

The final exam will consist of two sections. Section I will consist of several fill-in-the-blanks type questions. All questions are drawn out of the lecture slides. If you study the lecture slides, you should be able to answer these very easily. Most questions are indeed very straightforward. Completing this section should take no more then 30 minutes.

Note that I do not expect you to memorize any equations; however, you should know the definitions and description of each concept. You should also be able recognize an equation and state what it is used for.

• Population, sample, sample size, inferential / descriptive statistics, parameters, statistics;
• Measures of tendency in data: mean, median, skewness, percentiles, mode, variance, outlier, standard deviation, quartiles, lower / upper forth, fourth spread (interquartile distance);
• How to draw histograms / box plots;
• Definition and axioms of probability, disjoint (mutually exclusive) events, relative frequency;
• Fundamental principle of counting, permutations, combinations, law of total probability;
• Conditional probability and Bayes rule, prior and posterior probabilities;
• Continuous and discrete random variables, probability mass function versus probability density function, cumulative distribution function, expectation operation, mean and variance of random variables and of their functions;
• Definitions of binomial, hypergeometric, geometric, negative binomial and poisson distributions – what do the corresponding random variables describe?
• Calculating probabilities of the binomial, normal and t-distribution – being able to read distribution tables – probabilities corresponding to a given critical value or the critical value corresponding to a given probability - interpret pdf and cdf from these tables. How to standardize a normal distribution to obtain one with zero mean and unit variance? Normal approximation to binomial distribution;
• Joint and marginal probability distributions, dependence of random variables, covariance, correlation coefficient;
• Meaning of a random sample, independent and identically distributed random variables, normal approximation of the sampling distribution (sample mean), the central limit theorem;

The final exam will concentrate more on the following concepts

• Definition of confidence interval. What does it mean to have 100(1-\(\alpha\))% confidence interval? How to compute \(\alpha\) from confidence level, and obtaining the corresponding z-critical values from normal dist. tables;
• How to compute confidence interval for large samples – proportion of success and mean of a parameter? How to compute the interval for smaller sample sizes? Using the t-distribution tables;
• Hypothesis testing – meaning of null and alternative hypotheses – how to construct the hypotheses for a given problem. Definition of rejection region, type I and type II error.
• Different tests used for hypothesis and when to use which one – for known or unknown \(\sigma\), small or large sample size, the z-test and t-tests for population means and proportion of successes (success rate).
• Definition, meaning and importance of p-value, how it is computed.
• Tests involving comparing two populations – comparing population means and proportion of successes for small or large sample sizes, and for known or unknown \(\sigma\). Pooled t-test procedure, when and how to use them.
• When and how to use paired comparisons, paired t-tests. Confidence interval for $\mu_1 - \mu_2$ and $p_1 - p_2$.
• Tests involving more than two populations – ANOVA testing. Definition and meaning of factors and levels, one way or multiway ANOVA (computing multiway ANOVA is not included), sources of variation: within sample, between sample and total variations, mean square and sum for treatments and error. The meaning and definition of F-statistic, its relation to chi-square distribution, its computation from tables. (F-test for testing variances is not included).
• Interpretation and construction of an ANOVA table for one-way ANOVA.
• Multiple comparisons procedure – what to do if the null hypothesis for ANOVA is rejected. Simplified Tukey test and Q-statistic from studentized range distribution OR the simpler Fisher’s LSD test (from your text).

The second part of the exam will consist of problems from computing confidence intervals, hypothesis testing (one or two populations), and anova testing (with multiple comparison analysis).

Not that this is a closed book – closed notes exam, but you are allowed one equation sheet. You may use both sides of this sheet. IMPORTANT: Similar to the midterm exam, the equation sheet MAY NOT include any definitions, descriptions, or explanations. It may include only and only equations along with a very brief title for each equation. You MAY NOT write the names / descriptions of the variables used in equations. The equation sheet will be collected at the end of the exam. Make sure that you have your name on the equation sheet. Use the previously provided sample equation sheet as a guide on how to prepare one for this exam.

Again, note that the names, definitions, descriptions of variables may not be included in this equation sheet. The variable names are mostly standard in probability and statistics, and therefore you should know these. Furthermore, allowing definitions in equation sheets also opens the door for cheating in the fill-in-the-blank section. Please note that this policy will be strictly enforced. Calculators may be used in the exam; however equations may not be programmed into a handheld calculator or computer. Doing so will be considered as cheating and academic dishonesty. Also, you must show all work to get full credit. Numerical answers not accompanied by proper solutions will earn only 25% of the full mark for that question, even if they are correct.

Note that the project will count as 20-30% of your grade. The completeness of the report, the relevance of the application, the inherent difficulty in the test you have decided to conduct, the proper experimental design, well explained results, conclusions and interpretations, comparing different scenarios will all affect the final project grade.

Similarly, if you are caught cheating in this exam, your grade from the exam as well as from the project will be zero. It is also my responsibility to remind you that disciplinary action as governed by the university policies may also be pursued in all cases of academic dishonesty.

I will proctor the exam. Please feel free to let me know if you have any questions. The official exam duration is two hours. However, I will give additional time to those who need it. We can only do so if you come to the exam earlier, say at 7:30AM.

Good luck to all of you. I am sure you will all do very well, not only in this exam, but wherever and whenever you may need to use these statistical concepts.