Test 3 Information

Material Covered

Test 3 is cumulative but will emphasize material from sections and 4.1– 4.10. (Note: we have not covered Sections 3.5, 3.7 or 3.9.)

There is a summary at the end of each chapter that briefly reviews the most important topics of the chapter.

Format

The exam will include both an in-class portion and a take-home portion. For the in-class portion, no notes are allowed, but I will provide you with a version of the tables in the back cover of the book.

Some Additional Comments

- Be sure to look over your old homework so you can fix any problems detected there.
- No mystery numbers allowed. It should be clear where every number comes from.
  - If you use a calculator or computer to get a number, it must be clear from the work on your paper how someone else could get that number. (Write down the R code, for example.)
  - When doing combinatorics problems, make it clear where the component numbers are coming from.
  - Round as late as possible. Keep three significant digits. (Leading 0's are not significant digits.)
- Use notation well.
  - You are required to understand and use the notation we have introduced in class. This includes correct use of the equals sign (=).
  - If you received a "notation" comment on a problem set, be sure you understand it.
  - You may invent notation as long as you explain it.
- Don't be afraid to use words.

  In any case, do your work in "paragraph order" (left to right, top to bottom).
- When doing probability problems be sure to identify the events and random variables involved.

  For working with random variables, for example, it is often good to have statements like each of the following exaples as part of your solution:
  - Let X = the time until the next customer arives.
    (Describe the random variable in words.)
  - Then $X \sim \mathsf{Exp}(\lambda = 20)$.
    (Specify the particular distribution of the random variable if it is one of our familiar examples.)
  - $\mathrm{P}(X \geq 30) = 1 - \mathrm{P}(X \leq 30) =$ `1 - pexp(30,20)`
    (Identify the probability you are calculating and the R code used to get it.)

  A similar approach should be used for other probability problems as well. This approach will help you think clearly and avoid errors. It will also help me grade your work.

Things you should be sure to know

1. basics of random variables (continuous and discrete, including jointly distributed random variables)
   (a) pmfs, pdfs, cdfs, probability and quantile calculations
   (b) how and why to use the cdf method
   (c) mean, expected value, variance, covariance
   (d) conditional and marginal distributions, independence
   (e) favorite rules for combining random variables ("the expected value of a sum is the sum of expected values", etc.)
   (f) cdf method
2. favorite distributions (now including Chisq(), T(), and F())
   (a) situations where they are (or might be) good models
   (b) connections to sampling distribuitons and inference
   (c) R functions: `dnorm()`, `pnorm()`, `qnorm()`, `rnorm()`, etc.
3. moment generating functions, especially
   (a) definition and calculation
   (b) obtaining mean and variance from mgf
   (c) uniqueness and distribution recognition
   (d) mgf's for linear transformations of random variables
   (e) mgf's for sums of independent random variables
   (f) use in proof of central limit theorem
4. estimators and sampling distributions
   (a) lingo (estimator, estimate, and estimand, statistic, sampling distribution, standard error)
   (b) notation ($\overline{X}$ vs. μ vs. $\mathrm{E}(X)$ vs. $\boldsymbol{X}$; s vs. σ; $\hat{\theta}$; $\overset{\text{iid}}{\sim}$, etc.)
   (c) unbiased estimators (including why we define sample variance the way we do)
   (d) method of moments estimators
   (e) iid random sampling
5. quantile-quantile plots (especially normal-quantile plots)

6. inference procedures
   (a) 4 steps of a hypothesis test
   (b) meaning of p-value and confidence interval
   (c) power, type I error, type II error
   (d) coverage rate (nominal and actual)
   (e) duality of confidence intervals and p-values; pivotal quantities
   (f) creating new hypothesis tests
7. inference based on normal distributions
   (a) Central Limit Theorem
      i. statement
      ii. importance
      iii. use to obtain other limit theorems (e.g., binomial is approximately normal under certain conditions)
      iv. approximating binomial probabilities using the normal distribution (including continuity correction, and rule of thumb for use)
   (b) confidence intervals
      i. confidence level, coverage rate
   (c) hypothesis tests and confidence intervals for a mean
      i. differences between situation when σ is known (Fairy Land) vs. unknown
      ii. `t.test()`
      iii. duality of confidence intervals and p-values
      iv. paired tests
   (d) hypothesis tests and confidence intervals for a proportion
      i. `prop.test()` vs. `binom.test()`
      ii. ways to define (approximate) confidence intervals for a proportion
      iii. sign test (no ties version)
8. simulations
   (a) things simulations can be use for
   (b) designing a simulation for a given task
   (c) writing simulations in R
   (d) summarizing results with tables or graphs
9. linear algebra
   (a) notation for vectors
   (b) dot products and projections
   (c) sums of products as dot products, sums of squares as squared lengths
   (d) facts about distributions of dot products
   (e) decomposition of a data vector $\boldsymbol{X}$ into orthogonal components and distribution of S^2.
   (f) why we choose $\boldsymbol{u}_1 = \frac{\mathbf{1}}{|\mathbf{1}|}$
   (g) distribution of S^2, Chisq(n-1)-distribution, etc.