Material Covered

Test 1 covers Chapters 1–5 in the Notes. Also see the dope sheet for links to worksheets.

Format

The test will include a variety of problem formats. Possible formats include (but are not limited to)

- TRUE/FALSE
- multiple choice
- matching
- short answer
- mathematical calculations (of probabilities, etc.)
- constructing R commands to get something done
- interpreting numerical or graphical output from R.

A note on short answer questions: When I grade short answer questions I am looking for answers that are clear, concise, complete, and correct. Be sure your answer gets to the heart of the matter and avoids extraneous rambling. If asked for an example, be sure to choose a particularly clear example.

Things to be sure you review

This is not intended to be an exhaustive list, but the items below are important things to review.

1. Graphical and numerical summaries of data
    - mean, median, standard deviation, variance, interquartile range (IQR) and when to use them
    - how to calculate the above using R
    - histograms, density plots, boxplots, scatter plots, etc., when to use them, and how to interpret them
    - how to create these plots and numerical summaries in R
    - computing numerical or graphical summaries separately for each of two or more groups
2. Probability
    - terminology: outcome, event, sample space, random variable, mutually exclusive, independent
    - emprical vs theoretical probability (methods and relative advantages)
    - probability axioms and rules

- conditional probability and independence (keep independent and mutually exclusive straight)
- tree diagrams and how to use them (technically, this is optional, since all problems can be done without a diagram, but many students find this technique helpful)

3. Discrete and Continuous Random Variables

   - definition of pdf, pmf, cdf
   - probability tables for discrete random variables and how they related to the pmf
   - determining the pdf if given a kernel for a distribution
   - determining the pdf if given the cdf for a distribution
   - using a pmf or pdf to compute probabilities, expected values (means), and variances for a random variable (`antiD()` or `integrate()` can help with integrals when numerical results are desired.)
   - important families of distributions (see table in notes, plus the t-distributions)
   - using `dnorm()`, `pnorm()`, `qnorm()`, `rnorm()` and their cousins for other distributions.
   - fitting a distribution to data using method of moments (by hand) or maximum likelihood (using `fitdistr()` or `gf_fitdistr()`)
   - computing mean (expected value) and variance (and standard deviation) for a random variable
   - propagation of means and variances (and standard deviations) for linear combinations of independent random variables; the Pythagorean Theorem for Standard Deviations
   - using quantile-quantile plots to assess whether a probability model is a good fit to a data set.
   - simulating random variables (using `rflip()`, `sample()`, and `resample()`, and functions like `rnorm()`)

     Note: When a method is available that does not require simulation, that method is preferred. But if you cannot do something without simulations, using simulations is better than doing nothing.

4. Sampling distributions and confidence intervals

   - key terms: population, sample, parameter, statistic, sampling distribution, estimand, estimate, estimator, confidence level, critical value, standard error
   - the 68–95–95.7 Rule for normal distributions
   - other special properties of normal distributions
   - Central Limit Theorem
   - Calculation of confidence intervals for a mean from numerical summaries and from raw data (using `t.test()`)
   - confidence intervals for a proportion

Further suggestions

- 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 probability problems, make it clear where the component numbers are coming from.
  - Round as late as possible. Keep three significant digits in your final answer. (Leading 0's are not significant digits.) We'll talk more about how to think about significant digits statistically soon.
- 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 (=).
  - 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
  - Let $X =$ the number of free throws Freddie makes.
    (Describe the random variable in words.)
  - Then $X \sim \text{Binom}(20, 0.8)$.
    (Specify the particular distribution of the random variable if it is one of our familiar examples.)
  - $P(X \geqslant 16) = 1 - P(X \leqslant 15) = \ldots$
    (Identify the probability you are calculating and how are calculating it. Don't just write down the final number.)

  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.
- Don't neglect the "easy stuff".

  Some of the probability problems are challenging and require combining several ideas in order to get to a solution. Don't let the fact that some of these are challenging make you forget about easier things.
- Be sure you are solid on all the basic definitions and probability rules (intersection, union, complement, conditional probability, independence, mean, variance, standard deviation, etc.)