Material Covered

The test is cumulative, but it may slighly emphasize the newer material in chapters 6–9.

Format

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

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

  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.

  In short: Good answers are to the point and indicate that you understand what the point is.

On the exam you may use the following:

- RStudio (but no other software)
- Blank paper and something to write with. (You should submit your written work vie Gradescope immediately after the exam.)

The use of any other assistance (notes, internet, software, etc.) is not allowed.

Things to be sure you review

This is not intended to be an exhaustive list, but the items below are important things to review from the material since Test 1. See the Test 1 review sheet for earlier topics.

1. Sampling distributions and confidence intervals
    - key terms: population, sample, parameter, statistic, sampling distribution, estimand, estimate, estimator, confidence level, coverage rate, critical value, standard error
    - Central Limit Theorem
    - important template: estimate ± critical value · SE
    - interpreting confidence intervals (What does the 95% in a 95% confidence interval mean? 95% of what?)
    - specific situations, confidence intervals for
        - a mean
        - a proportion
        - the difference between two means or two proportions
        - regression coefficients
    - obtaining critical values via `qt()` or `qnorm()`
    - automating using `t.test()`, `lm()`, `prop.test()`, `binom.test()`, `confint()`.
2. Simulating random variables (using `rnorm()` and its cousins)

    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.
3. Means and variances (and standard deviations) for linear combinations of (independent) random variables; the Pythagorean Theorem for Standard Deviation
4. Normal and t-distributions
    - the 68–95–99.7 Rule
    - other special properties of normal distributions
    - how normal and t-distribtions compare
    - R functions for each
5. Uncertainty
    - what uncertainty is and why it is important
        - standard uncertainty, absolute uncertainty, relative uncertainty
    - estimating uncertainty without data (using distributional assumptions)
    - estimating uncertainty using data (standard error)
        - important standard error formulas ($\frac{s}{\sqrt{n}}$, etc.)
    - delta method for estimating uncertainty (propagation of uncertainty)
        - main formula for propagation of error (assuming independence):

$$W = f(X, Y) \Rightarrow u_W \approx \sqrt{\left(\frac{\partial f}{\partial X}\right)^2 u_X^2 + \left(\frac{\partial f}{\partial Y}\right)^2 u_Y^2}$$

and what it means

  - versions that have only one variable or more than two variables
- Relative uncertainty, including the nice rule for products and quotients.
- Using uncertainty to determine how many digits to report.

6. Linear models

- algebraic form (equation with β_i's in it)
- error term and residuals (don't conflate errors with residuals)
  - method of least squares
  - the four assumptions of a linear model (LINE)
  - using residuals (residual plots, qq-plots of residuals, acf plots of residuals) to check the assumptions of the linear model. [`plot()`, `mplot()`, `gf_qq()`, `resid()`]
- how to fit linear models using `lm()` and interpret R's regression output
  - using `I()` on the right side to avoid special interpretation of things like `+ * ^`, etc.
  - getting more information using `summary()` or `msummary()`
  - locating estimates for coefficients (slope and intercept) in ouput
  - locating standard errors (uncertainty) in output
  - locating $\hat{\sigma}$ in output
  - using `confint()` to get confidence intervals for the parameters.
  - "residual standard error" and what it estimates ($s = \mathsf{RMSE} = \sqrt{\mathsf{MSE}} = \sqrt{\mathsf{SSE}/\mathsf{DFE}}$)
  - R (correlation coefficient) and R^2 (coefficient of determination) and how to interpret them
  - the relationship between degrees of freedom and sample size
- interpreting model parameters
- converting back and forth between model equations and R's notation for a model
  - example: `lm( y ~ x )` corresponds to $Y = \beta_0 + \beta_1 x + \varepsilon$ if both explanatory and response are quantitative
- confidence intervals for model parameters.
  - You do NOT need to know formulas for standard errors of linear model parameters, just how to find them in R output and interpret them.
- computing confidence intervals and prediction intervals for the response in a linear model (and the difference between them)
- using transformations to convert non-linear relationships into a model that can be fit with `lm()`.
- `makeFun()`

7. Hypothesis tests

- four step approach to hypothesis testing
- interpreting p-values
- the duality between p-values and confidence intervals
- important template: test statistic $= \dfrac{\text{data value} = \text{hypothesis value}}{\mathsf{SE}}$
- obtaining p-values via `pt()`
- automating using `t.test()` or `lm()`.

8. Nonlinear models with `nls()`
    - How to specify the model formula
    - Use of `start` to specify paramters and initial values for the search
    - Interpretting output of `summary()`, `msummary()`
    - How models fit with `nls()` and `lm()` might differ

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.
- 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).