

22S:166 Computing in Statistics
PRACTICE PROBLEMS for FINAL EXAM 2007
Fall 2006
Final Exam

Copy your code and output for each problem into a Microsoft Word document or a \LaTeX document. **Include your name and a 4-letter or 4-digit code for me to use in posting your grade. The code must NOT be part of your Soc Sec number or student i.d.**

Submit your exam by uploading your document into ICON.

1. A university bookstore wants to store information about the textbooks that it sells and in which courses each one is used. Some courses require more than one textbook, and the same textbook can be used in more than one course. Here's the way the bookstore manager set up a single file to hold the information. There was a problem when a couple of courses required 3 or more textbooks.

- course number (e.g. 22S:166)
- course section (e.g. Section 1)
- course semester (e.g. Fall 2006)
- course textbook1 ISBN number (this is a real-world unique identifier for books)
- course textbook1 title
- course textbook1 first author
- course textbook1 publisher
- course textbook1 copyright date
- course textbook1 current selling price
- course textbook2 ISBN number
- course textbook2 title
- course textbook2 first author
- course textbook2 publisher
- course textbook2 copyright date
- course textbook2 current selling price

Lay out a relational file structure for storing these data items. Put it in 3rd normal form.

For this problem, type your answer and include it in the document you send to me.

2. The `stats` package in R includes two functions for carrying out a two-sided hypothesis test as to whether a population proportion p is equal to some specified value p_0 , that is:

$$H_0 : p = p_0$$
$$H_A : p \neq p_0$$

The `prop.test` function uses a Chi-square approximation for the hypothesis. The approximation may be computed *with* or *without* the Yates continuity correction. See the function documentation for how to specify your choice.

The `binom.test` function calculates the exact p-value for the hypothesis test. It would be expected to give more accurate results with small samples than either of the approximate tests.

- (a) Use the `rbinom` function to simulate the number of successes in one binomial sample of size 16 when the true success probability is 0.5. Turn in your code and output. (The output should be a single number.)
- (b) Use the `binom.test` function to calculate the p-value of a two-sided test of the null hypothesis that $p = 0.5$, using your simulated data as the data. Turn in your code and the output.
- (c) If we were conducting our hypothesis test at significance level .05, we would reject H_0 if the p-value were less than .05 and fail to reject otherwise. Which decision results from your results in the previous step? (It could have turned out either way, incidentally.) Write one sentence answering this question.
- (d) Recall that the `size` of a statistical test is the probability of rejecting the null hypothesis when the null is actually true. Carry out a simulation study comparing the size of two-sided hypothesis tests regarding a binomial success probability carried out at a nominal significance level of .05 by the three different methods:
 - i. Chi square test, Yates continuity correction not used
 - ii. Chi square test, Yates continuity correction used
 - iii. exact binomial testunder the conditions
 - i. sample size = 16
 - ii. true success probability = 0.5

While testing your simulation procedure, use a very small number of replicate datasets to save time. When doing your final run, use 2000 datasets.

Turn in the code for your simulation study and a table showing the results. Include estimates of the standard errors of your estimates of size.