Biost/Stat 578A
Emerson, Spr 13
May 6, 2013

Homework #3: Written problems to be submitted electronically by 11:59 pm on Sunday, May 12.

The following setting applies to all problems.

Consider a setting in which we desire to design a clinical trial to test a new experimental drug for versus placebo in a randomized, double blind clinical trial. The primary endpoint is treatment response (a binary endpoint) following 28 days of treatment. With placebo, the proportion of patients evidencing response is presumed to be 20%, and it is judged that a response rate of 30% would be clinically important. (Note that this "design" hypothesis corresponds to an absolute improvement of 10% in response rates or a 50% relative improvement in response rates.) We intend to randomize patients in a ratio of 1 experimental : 1 placebo. We desire a one-sided level of significance of 0.025.

In the problems that follow, you may want to make use of the RCTdesign functions

- `seqDesign()` to create a RCT design,
- `seqOC()` to obtain operating characteristics such as power, average sample size, and stopping probabilities,
- `seqInference()` to display the inference corresponding to stopping boundaries,
- `plot()` (or equivalently `seqPlotBoundary()`), `seqPlotPower()`, `seqPlotASN()`, and `seqPlotStopProb()` to produce plots displaying or comparing RCT design characteristics, and
- `seqEvaluate()` to obtain standardized tables and plots of operating characteristics for a design.

1. Consider first a fixed sample design (i.e., a design with only a single analysis). For this problem perform all calculations using variance estimates based on the null hypothesis. (This can be effected using the `seqDesign` argument `variance="null"`.)
   a. What sample sizes are required in order to be able to detect the hypothesized "design" alternative with 80%, 90%, 95%, and 97.5% power?
   b. If we choose to use a sample size of 1000 subjects, what will be our statistical power to detect the "design" alternative?
   c. If we choose to use a sample size of 1000 subjects, what will the the alternative treatment effects that will be detected with 80%, 90%, 95%, and 97.5% power?
2. Repeat problem 1, but use variance estimates based on the alternative hypothesis. (This can be effected using the `seqDesign` argument `variance="alternative"`.)
3. Repeat problem 2, but now consider the hypothesis that the treatment would provide a 10% absolute improvement above a placebo response rate of 35%.
4. Repeat problem 2, but now consider the hypothesis that the treatment would provide a 50% relative improvement above a placebo response rate of 10%.
5. Consider again the setting of problem 2 (with the original "design" hypothesis), but now suppose that we desire to use a group sequential design having a maximum of 4 analyses (i.e., 3 interim analyses, and one final analysis) using O'Brien-Fleming boundary relationships (P = 1, R = 0, A=0 within the unified family) for any early stopping decisions.
   a. Suppose that we only want to stop early for decisions of efficacy and that the analyses are to be equally spaced in terms of statistical information. What would be the maximal sample size if we maintain power of 97.5%? What would be the power to detect the "design" alternative if we maintain the sample size that was found to provide 97.5% power in problem 2a?
   b. Suppose that we only want to stop early for decisions of futility and that the analyses are to be

equally spaced in terms of statistical information. What would be the maximal sample size if we maintain power of 97.5%? What would be the power to detect the "design" alternative if we maintain the sample size that was found to provide 97.5% power in problem 2a?

c. Suppose that we only want to stop early for decisions of both efficacy and futility and that the analyses are to be equally spaced in terms of statistical information. What would be the maximal sample size if we maintain power of 97.5%? What would be the power to detect the "design" alternative if we maintain the sample size that was found to provide 97.5% power in problem 2a?

d. Repeat part c, except now presume that the first analysis should occur very early as might be required by regulatory agencies when limited phase 2 data are available. Hence, assume that the four analyses are to be performed at times corresponding to 10%, 50%, 75%, and 100% of the planned total statistical information.

e. Repeat part c, except now presume that the first analysis should only occur after 40% of the planned total statistical information is available (this might be necessary to satisfy the regulatory requirements of adequate safety data). Hence, assume that the four analyses are to be performed at times corresponding to 40%, 50%, 75%, and 100% of the planned total statistical information.

6. Repeat problem 5, but use Pocock boundary relationships (P = 0.5, R = 0, A=0 within the unified family) for any early stopping decisions.

7. Provide a full comparison of the stopping rules obtained in problems 5c and 6c when maintaining power at 97.5%. Be sure to comment directly on comparisons of the power curves, maximal sample size requirements, ASN curves, stopping probabilities as θ varies.

8. Suppose now that we want to use designs that are efficient (on average) within the one-sided symmetric boundaries derived having type 1 error of 0.025, power of 0.975 to detect the design alternative of 30% response rate vs a placebo response rate of 20%, and the same boundary shape function for both efficacy and futility boundaries. Consider the spectrum of boundary shape functions within the Wang and Tsiatis boundary family (so as a function of P > 0 with A=0 and R=0 in the unified design family).

a.. What RCT design in this family would have the lowest ASN under the null hypothesis? under the alternative hypothesis?

b. What RCT design in this family would have the lowest ASN under the intermediate hypothesis of an absolute difference of 0.05 in response rates?

b. What RCT design in this family would have the lowest ASN under the intermediate hypothesis of an absolute difference of 0.075 in response rates?

9. Repeat problem 8, but consider the Kim and DeMets power family of error spending functions (so as a function of P < 0 with `design.family="E"`).

10. Repeat problem 9, but consider the Hwang, Shih, and DeCani family of error spending functions (so as a function of P with `design.family="Hwang"`).

11. Explicitly compare the designs found in each part of problems 8, 9, and 10, and comment on any substantial differences among the "optimal" designs.