Course EPIB-681: Data Analysis II [Winter 2004]

Assignment 1

material in www.epi.mcgill.ca/hanley/c607/ unless otherwise specified
(username: c607 ; password: 8 letters, H***J*## both case-sensitive)

1 You observe 3 "positives" in a simple random sample of size n = 20 from a certain 'source'.
- Calculate the usual large-sample 95% (frequentist) CI for the proportion positive (π) in the source [do not use a continuity correction]. Comment.
- Obtain the 'exact' 95% (frequentist) first-principles Klopper-Pearson CI for π. [see p6 of JH's Notes for Chapter 6, and p3 of Notes for Chapter 8.1; you can obtain it (i) using interpolation, using a table such as Binomial Table in Moore and McCabe; *or* (ii) by trial and error using e.g. the Excel spreadsheet in Resources for Ch 6 or Ch 8, or by calculating the binomial tail areas manually.
- Calculate the 'approximate' 95% (frequentist) 'Wilson' CI for π. [see p5 of JH Notes for Chapter 8.1 in course 607].
- Obtain/calculate a 'logit-based' and a 'log-based' 95% (frequentist) CI for π. [see p6 of JH Notes for Chapter 8.1 in course 607]. You may obtain them manually, or via 'binomial regression' software [generalized linear model].
- How close are the *upper* limits obtained by the 4 approximate methods? the *lower* limits? Which limit (upper or lower) is more sensitive to the various approximations, and why?

2 You observe 0 "positives" in n = 41 'cases' { e.g. 0/41 seropositives in HPV rct had been vaccinated. The *parameter* of interest is the corresponding proportion π in an effectively-infinite-size trial.}
- Which of the above (frequentist) methods yield sensible answers for the upper limit of π?
- If your prior beliefs about π can be represented by a Beta(a=1,b=9) probability distribution, what is the corresponding (post-data) distribution for π? Sketch this posterior distribution [if you have not taken course 607, consult someone who has. If you do not know First Bayes, you can use the "Bayesian Inference for a Proportion (Excel)" spreadsheet under Other resources for Chapter 8 in the c607 website:- specify the prior using a parametric curve option [start at row 32 and choose option 2]

3 Patients undergoing colorectal surgery were randomly assigned to routine intraoperative thermal care or additional warming. In a double-blind protocol, their wounds were evaluated daily until discharge from the hospital and in the clinic after two weeks. Surgical-wound infections were found in 18 of 96 patients assigned to hypothermia (19 percent) but in only 6 of 104 patients assigned to normothermia (6 percent, P=0.009).
- Calculate a point- and a 95% interval- estimate for the *difference* in infection rates.
- Convert the point estimate of the difference into the "number required to treat" via the formula 1/(Infection Rate if do not treat – Infection Rate if treat)

  [if 19/100 would develop infection without intervention, and 6/100 despite it, then intervening on 100 prevents 19 - 6 = 13 , i.e.. need to intervene on approx 8 (i.e. 100/13) to prevent 1 infection]
- Convert the upper and lower 95% limits for the rate difference into the corresponding limits for the number required to treat.

  [taken from Q -15- in Homegrown Exercises around Ch. 8, course 607: article N Engl J Med 1996;334:1209-15. Note that in this example, JH is using the term 'rate' as a 'proportion-type' rate -- just as in lay language, we talk of a complication 'rate', (a risk, or cumulative incidence, up until the clinic visit) and of the unemployment 'rate' (a prevalence)]
- Calculate point- and a 95% interval- estimates for the relative risk and the odds ratio.
- Stata's notes state "We recommend that test-based confidence intervals be used only for pedagogical purposes and never be used for research". Practically, how do test-based CI's fare in this example? Consider all 3 effect measures, RD, RR and OR .. *see pp5-6 of Notes for Ch 8.2*

4 [From "Questions" on pp142 Chapter 7 of Rothman 2002] Q3 (prev. & risk data) and Q4 (risk vs c-c data)