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

Assignment 2

`material in www.epi.mcgill.ca/hanley/c626/  unless otherwise specified`

`( username: c607 ; password: 8 letters, H***J*##  both case-sensitive )material in`

1 Which is the *single biggest* flaw in the analysis of the scouting injuries [below]. List two others that on their own might be major -- but not nearly as large as the distortion produced by *the big one* !

2 Question 1 from "Questions" on p 142 at end of Chapter 7 of Rothman2002 (denominator errors)

3 Assuming (i) no temporal trend (ii) Poisson variation from year to year, and (iii) the fact that a sum of independent Poisson random variables is again a Poisson random variable, obtain a 95% CI for the average number of hurricanes with winds greater than 110 mph to strike the USA. Express these using time units of (a) a year (b) a decade (c) a period of 95 years (same length as period reported) and (d) a century. Reference your methods, and show the steps. [if wish to use computer, see hurricane data/program(s) under **Datasets**]

4 Calculate a 95% CI for the SIR and perform a test (at alpha = 0.05 2-sided) of SIR=1 for the Alberta Sour Gas Study below. Restrict attention to the 33 vs. 36.3 [Index Area 1970 Cohort Females vs. (1) Southern Alberta excl. Calgary, Lethbridge, & Medicine Hat (RP1)]. Describe your procedures/ steps.

Carry out the same inference procedures, but imagine the concerned area or cohort was much smaller .. so that 3 cases were observed where 0.45 were expected. Again, describe your procedures/ steps. *[if you do not have access to a computer/calculator, you could use tables from JH's Notes on Poisson]*

Rothman (p 142) was concerned about errors in the *denominators* of rates. How about the effect of sampling variation in the reference *rates* [applied to denominators to arrive at 'expected' numbers']?

5 USA Today stated that "*The number of hurricanes with winds greater than 110 mph has declined since the 1950s*" [cf. Question 2].

Contrast the rate in the last 4.5 decades (index category) with that in the first 5 decades of the century (reference category), Calculate 95% CI's for the rate difference and the rate ratio, respectively.

Comment on the USA Today statement.

List two other methods for arriving at a CI for the Rate Ratio [calculations not required, but give references to similar examples in Notes or Textbooks, and 'line up' the 'inputs' to the corresponding formulae].

6 [OPTIONAL] Refer to the data from John Snow's study, given on bottom of column 3 of page 1 of handout for Sept. 05 lecture for Med2 [on med2 website, reachable from link at top of 626 website: username med2, password: same as for the cxxx epidemiology courses]. Calculate a 95% CI to accompany the rate ratio of 13.3. Do the same for the ratio estimates based on the denominator series of 100 and 1000 (first column, page 2... [in practice, you would not observe the quasi-denominators shown there, but rather these expected numbers ± some sampling variation]. Why are the CI's based on the 100 or 1000 *wider* than the one based on the actual "return which was made to Parliament"?

7 [OPTIONAL] Refer to the article A POPULATION-BASED STUDY OF MEASLES, MUMPS, AND RUBELLA VACCINATION AND AUTISM [under **Nov. 11 lecture in med2,** reachable from link at top of 626 website: username med2, password: same as for the cxxx epidemiology courses]. See also page 2 of the handout of Nov. 11, dealing with CI's for ratios, and the fact that for ***log-based CI's*** (instead of the usual ± a margin of error for 'regular' statistics) for ***ratios***, one can calculate a "multiplied-by/divided-by" factor in order to arrive at the upper/lower limits. To convince yourself, calculate the CI's for the ratio of 13.1 on page 2, and the 1.44 ratio on page 3, by your usual way, and compare them with the answers from the "multiplied-by/divided-by" method shown.

8 The following questions concern the article A CONTROLLED TRIAL OF A HUMAN PAPILLOMAVIRUS TYPE 16 VACCINE and are adapted from question 2 in the December 2002 exam in course 626. The abstract is given below; if needed, the full article can be reached via the link on the course 626 webpage. The conditional approach in this question is *similar to that behind the worked e.g. on page 29 of JH's Notes on Poisson Distribution*]

a. The study employed a fixed-numbers of events design" (1st sentence Statistical Analysis section).

Why this design rather than a "fixed number of woman-years-of-follow-up" design?

b. With I denoting incidence, v denoting the vaccinated and u the unvaccinated, Efficacy (E) is defined here as a percentage

$$E = 100 \times (I_u - Iv) / I_u = 100 \times (1 - I_v / I_u)$$

Consider a very large R.C.T. (so random variation is not an issue), with 1/2 receiving the vaccine and 1/2 the placebo, and concentrate on the ***total number of cases*** (of persistent infection).

What is the relation between the proportion (P) of these cases that would be in the vaccinated group (i.e. what fraction of cases would be 'exposed' cases) and the vaccine efficacy E? To answer, calculate for every 1 case in the unvaccinated, how many cases there would be in the vaccinated; then express the #v as a proportion of (#u + #v).

E (%)	0	25	50	75	80	90
# u cases	1	1	1	1	1	1
# v cases						
	___	___	___	___	___	___
P=proportion of cases that received v	___	___	___	___	___	___

c. Suppose that in the actual (finite) study, subject as it was to random variations, the authors had analyzed the data when the total number of cases was 31, i.e. when the observed proportion of cases that had been vaccinated was p=0/31 i.e., when the point estimate for P was p=0.0. This point estimate translates into an 'exact' 95% 2-sided CI for P of 0.0 to 0.11 [*binomial-based*, *see Ch8*].

From this CI, and interpolation in the table you constructed in part b, find 'exact' 95% limits for E.

9 The large-sample methods for obtaining a CI for a rate ratio are accurate when there are enough events in each of the compared categories,. But in Q8 above and in the "Women are Safer Pilots" example below, the small number of events in one of the categories renders large-sample methods inaccurate or even impossible. In such situations, the *conditional* approach, in which one bases the inference on the distribution of the number of events in one category, *conditional* on the *sum* of the numbers of events in the two categories, is a way around this problem (we use a similar strategy when we employ Fisher's exact test).

Suppose we are interested in the rate of accidents in women relative to men (i.e. the rate *ratio*). Assume that on average, the women pilots fly *just as many* hours as the men pilots, and that all other relevant factors are equal [although they probably are not!]. Show how you would, based on the information given, calculate a CI for the rate ratio [detailed calculations not required, but show key points]
Repeat the steps, but now assume that on average the women pilots fly *half as many* hours as the men.

Is Scouting Safe?

Over the past year, leaders have been showing a growing commitment to provide each member a safe and enjoyable Scouting experience. In support of efforts in the field, we conducted a study to establish baseline data on scouting accident and injury trends so that we can make informed decisions about activity precautions or the need for higher safety standards. This column highlights the findings The first question we asked ourselves was, "Is Scouting a safe program for members?"

Statistics Canada, Health Division, told us that 11 out of every 1,000 males aged 5-19 are hospitalized for at least one night a year. When we compared similar information taken from Scouting accident forms, we found our members are hospitalized at a rate of only one per thousand a year. Given that we run active programs and heavily use the outdoors, Scouting falls far below the average rate for daily living risk to males in this age group.

Having established this, let's look at the main kinds of accidents and injuries that do happen to Scouting members. Our study identified the types of injures that happened during the course of a normal Scouting year, excluding summer camps (Chart A). It also recorded the types of activities associated with the injuries (Chart B). When we examine the two sets of information, we begin to see some relationships...

The (Scouting) Leader (magazine) June/July 1991

A CONTROLLED TRIAL OF A HUMAN PAPILLOMAVIRUS TYPE 16 VACCINE

Background Approximately 20 percent of adults become infected with human papillomavirus type 16 (HPV-16).Although most infections are benign, some progress to anogenital cancer. A vaccine that reduces the incidence of HPV-16 infection may provide important public health benefits.

Methods In this double-blind study, we randomly assigned 2392 young women (defined as females 16 to 23 years of age) to receive three doses of placebo or HPV-16 virus-like –particle vaccine (40 µg per dose), at day 0, month 2, and month 6. Genital samples to test for HPV-16 DNA were obtained at enrollment, one month after the third vaccination, and every six months thereafter. Women were referred for colposcopy according to a protocol. Biopsy tissue was evaluated for cervical intraepithelial neoplasia and analyzed for HPV-16 DNA with use of the polymerase chain reaction. The primary end point was persistent 16 infection, defined as the detection of HPV-16 DNA in samples obtained at two or more visits. The primary analysis was limited to women who were negative for HPV-16 DNA and HPV-16 antibodies at enrollment and HPV-16 DNA at month 7.

Results The women were followed for a median of 17.4 months after completing the vaccination regimen. The incidence of persistent HPV-16 infection was 3.8 per 100 woman-years at risk in the placebo group and 0 per 100 woman-years at risk in the vaccine group (100 percent efficacy; 95 percent confidence interval, 90 to 100; $P<0.001$).All nine cases of HPV-16–related cervical intraepithelial neoplasia occurred among the placebo recipients.

Conclusions Administration of this HPV-16 vaccine the incidence of both HPV-16 infection and HPV-16–related cervical intraepithelial neoplasia. Immunizing HPV-16–negative women may eventually reduce the incidence of cervical cancer.(N Engl J Med 2002;347:1645-51.)

WOMEN ARE SAFER PILOTS: STUDY

LONDON- Initial results of a study by Britain's Civil Aviation Authority shows that women behind the controls of a plane might be safer than men. The study shows that male pilots in general aviation arc more likely to have accidents than female pilots. Only 6 per cent of Britain's general aviation pilots are women. According to the aviation magazine Flight International, there have been 138 fatal accidents in general aviation in the last 10 years, and only two involved women - less than 1.5 per cent of the total.

Woman News, page F1 The Montreal Gazette, August 21st, 1995

HURRICANES

USA TODAY, August 1995*. The number of hurricanes with winds greater than 110 mph has declined since the 1950s. Numbers of major hurricanes to hit the USA each decade:

1900s: **6**; 1910s: **8**; '20s: **5**; '30s: **8**; '40s: **8**; '50s: ***9***; '60s: ***6***; '70s: ***4***; '80s: ***6***; '90s: ***2***

[* The 1995 'season' is excluded; i.e. , the 1990s data are from the 5 seasons 1990, 1991, 1992, 1993 and 1994]

Table 15.1.4/1 from Alberta Sour Gas Study (Spitzer et al. 1985)

Standardized Incidence Ratios (SIR) excluding non-melanotic skin cancer for the index area 1970 cohort relative to three reference populations

Population	Observed	Expected (1)	Expected (2)	Expected (3)
Index Area 1970 Cohort *Males*	45	49.6	49.6	52.7
Index Area 1970 Cohort *Females*	33	36.3	34.2	35.1
Index Area 1970 Cohort Total	78	85.9	83.8	87.8
SIR		0.91	0.93	0.89
95% CI		(0.67 to 1.26)	0.68 to 1.29	0.65 to 1.23

(1) Southern Alberta excl. Calgary, Lethbridge, & Medicine Hat (RP1)

(2) Census Division 6 excl. Southern Alberta excl. Calgary (RP2) - contains DCCI

(3) Census Division 2 excl. Lethbridge (RP3) - contains SR