1. You observe 3 "positives" (e.g., left-handers) in a simple random sample of size $n = 20$ from a certain 'source.' Calculate, or otherwise obtain, a 95% (frequentist) CI for the proportion positive in the source ...

i. using usual large-sample approximation (untransformed) [do not use a continuity correction].

ii. using the 'exact' 95% first-principles Klopper-Pearson method. [explained on p6 of JH's 607 notes for Chapter 6; you can obtain it from the table handed out in class, which is also available on the 607 Resources Ch 8; or directly, or by trial and error, using the `Excel` spreadsheet in the same Resources; or using `R`.]

iii. using the 'Wilson' method. [formula given in Rothman2002 p132; jh's `R` code for this is in resources for proportions in website for course 634].

iv. manually, using the logit-transform [as per 1st page of "Statistical models for inference re epidemiologic parameters" handed out in class: if this is too cryptic, see p6 of Notes for Chapter 8.1 in webpage for course 607].

v. using logit link in 'binomial regression' [generalized linear model] [1]

2. Refer to "Competing risks of mortality with marathons" e.g. 1/31

Sudden cardiac death among participants

i. The authors report a "rate" of sudden cardiac death of 0.8 per 100,000 participants. As per EPIB601 terminology, which type of measure of (disease) frequency is this 0.8?

ii. Using a Binomial statistical model, calculate a 95% CI to accompany the 0.8. [In deciding which of the methods used in question 1 to use here, note that the journal limits you to reporting 1 decimal place.

iii. Use a Poisson statistical model to calculate a 95% CI to accompany the 0.8. Why does this CI agree so closely with the Binomial one *in this instance*? [*hint*: cf Contents: item 4. How does this distribution arise? in the webpage `http://en.wikipedia.org/wiki/Poisson_distribution`]

iv. Express the "rate" (and its corresponding CI) as the numbers of sudden cardiac death per million participant-hours of exercise.

Motor vehicle fatalities [from Results in fulltext]

"A total of 85 individuals died in fatal crashes on the marathon days in counties inside the course during hours when roads were closed. In contrast, 262 individuals died in fatal crashes on the control days in the corresponding counties and hours. Given that each marathon was paired with two control days, the discrepancy between observed and expected crash deaths on marathon days corresponded to a 35% relative decrease in risk (17% to 49%). This discrepancy was equal to an absolute decrease of 46 total crash deaths over the study (P<0.001)."

i. Obtain the actual P-value in two ways: (a) from a 1-sample z test of $\hat{p} = 85/347$ vs. $p_{null} = 1/3$ (b) from a 1-sample X^2 test of the observed numbers 85 and 282 vs. their null expectations (omit the continuity correction from both sets of calculations).

ii. The point-estimate of the percentage decrease was 35%. You can write the fractional decrease, i.e., the 0.35, as $1 - \widehat{IDR}$, where

$$\widehat{IDR} = \frac{85 \text{ deaths } /(750 \times 12 \text{ hours})}{262 \text{ deaths } /(750 \times 2 \times 12 \text{ hours})} = \frac{85/347}{262/347} \times \frac{2}{1} = \frac{\hat{p}}{1-\hat{p}} \times \frac{2}{1}.$$

From the observed $\hat{p}$ obtain a binomial-based 95%CI, $\{p_L, p_U\}$, for the corresponding theoretical value p. Substitute these limits into the expression for p to get

$$IDR_L = \frac{p_L}{1-p_L} \times \frac{2}{1}; \quad IDR_U = \frac{p_U}{1-p_U} \times \frac{2}{1}.$$

Do these limits for IDR agree with the 17% and 49% reported by the authors?

Comment: It is not clear if this is how the authors obtained a CI for the IDR. Another possible way would have been to use the SE[$\log \widehat{IDR}$] in a log-based CI for IDR[2]. The authors did use the 1-sample binomial for the P-value: in the Statistical Analysis section they say "Summary statistics were based on *binomial tests* and not adjusted for clustering."

---

[1] If using `R` ...

```
y=c(rep(1,3),rep(0,17)); y; constant=rep(1,20); constant;
summary( glm(y ~ -1 + constant,family=binomial) ); #check SE; sqrt(1/3 + 1/17);
```

if using `SAS` or `Stata`... see p6 of Notes for Chapter 8.1 in webpage for course 607.

[2] formula 7-5 in Rothman2002 p137, and limits for IDR calculated on p138; the same approach is described in lines [13] and [17] in the second column of page 2 of jh handout "Statistical models for inference re epidemiologic parameters."

Conditioning on the total of the two numerators is a common strategy in this type of statistical analyses – it converts a more difficult 2-sample problem (with 2 separate parameters ID_1 and ID_0) into a simpler problem involving a single random variable, ($Binomial[347, p]$), where the single parameter p is a function of the (comparative) *parameter of direct interest*, namely ID$\underline{\text{R}}$.

Whereas we had the option of staying with the more obvious 2-sample approach used in the log-based CI, there are several real examples where at least one of the two components of $\widehat{IDR}$ is – by itself – extreme and thus unstable. This applies not just for tests, but also, and more importantly, for CIs.

The "pilots" story (next) illustrates the benefits of a conditional approach; the vaccine efficacy study shows just how important this strategy is.

3. Refer again to the "Women are Safer Pilots" story (e.g. 4/31).

Suppose we are interested in the rate of accidents in women (index category, subscript $_1$) relative to men (reference category, subscript $_0$). Because in this example we are only given the *relative* sizes of the 'denominators', we are limited to estimates of the incidence density *ratio* (IDR).

In assignment 1, you might have used the CI template for the IDR_L from the bottom of page 138 of Rothman2002, and likewise for IDR_U. But a reviewer of your journal manuscript or thesis who saw that one of your numerators was only 2 might object that you cannot use a normal approximation, even with the log transformation, when the numbers of cases in one of the cells is *this* low. This reviewer might well ask that you use *exact* methods, and refer you to 'adult' rather than 'baby' Rothman, i.e., to Rothman1986.

The exact method Rothman describes there is based on the distribution of *one of the numerators, conditional on the sum of the two numerators*[3], and on the same Binomial inference that we use for a single sample – but now, with the sum of the two *numerators* as the 'n'. We derived the 'conditional' CI heuristically in the marathon study, but we repeat the derivation again here. It can be shown algebraicly [4] that when the person-time 'denominators' are PT_1 and PT_0, we can write the (*theoretical*) IDR *parameter* as

$$IDR = \frac{p}{1-p} \times \frac{PT_0}{PT_1}$$

where p is the *expected* fraction of the total number of cases that would arise from the index category($_1$) and $1-p$ is the expected fraction of the total number of cases that would arise from the reference ($_0$) category. Clearly, the fraction is an amalgam of the IDR and the relative sizes of the denominators: if both denominators were approximately the same size (as in an rct with 1:1 randomization), then if 2/3rds of the cases were in the exposed category, and 1/3rd in the unexposed, it would indicate that the IDR is 2. In an rct with say 3:1 randomization to the exposed and unexposed categories, and an IDR of 0.5, one would expect $3 \times 0.5 = 1.5$ exposed cases for every $1 \times 1 = 1$ unexposed case, i.e., an expected fraction (proportion) $p = (1.5/2.5) = 0.6$, or 60%, of the total number of cases would arise from the exposed category.

Empirically, in our example, of the 'n' =138 cases, the proportion $\hat{p}$ of the 138 that arose from the index category was only 2/138. This small fraction is not that surprising, given the small fraction of the total experience contributed by those in the index category. So it makes sense that the IDR estimate should be corrected for the difference in the sizes of the denominators, i.e.,

$$\widehat{IDR} = \frac{2/138}{1-2/138} \times \frac{PT_0}{PT_1}$$

i. The smaller the total number of cases, the noisier (less reliable) is the empirical fraction $\hat{p}$, and thus $\widehat{IDR}$. To reflect this, treat the $\hat{p} = 2/138$ as a binomial-based proportion with $'n' = 138$, and obtain an exact 95% binomial CI, $\{p_L, p_U\}$, for p. Since the table of exact binomial CIs handed out in class stops at $y = 2$; $n = 100$, to save you time, take it that the exact CI for p, based on 2/138, is –via the spreadsheet – $\{0.00176, 0.05137\}$. [The upper limit of the (approx. and much simpler) Wilson CI is quite similar: 0.05130, but the lower one is a bit different: 0.00398.][5]

ii. Assume that on average, the women pilots fly just as many hours as the men pilots (i.e. that $\frac{PT_0}{PT_1} = \frac{0.94}{0.06}$, and that all other relevant factors are equal [although they probably are not!]. Insert this ratio, along with $\hat{p}$, into the equation above to calculate $\widehat{IDR}$. Then use the same denominator ratio, but the exact lower limit p_L, to obtain a lower limit IDR_L for the IDR. Do the corresponding calculation to obtain IDR_U.

iii. Does $\{IDR_L, IDR_U\}$ agree with the P-value obtained by testing the observed proportion $\hat{p}$ (or observed number 2) against the null $H_0 : p = 0.06$? i.e. calculate $Prob[\leq 2 \mid \text{Binomial}(n = 138, p_{null} = 0.06)\ ]$.

iv. Repeat (ii)-(iii), assuming women pilots fly half as many hours as men.

[3] the same conditinal approach as is used in Fisher's exact test

[4] Derivation...
Denote the expected numbers of cases by $E[c_1] = ID_1 \times PT_1$ & $E[c_0] = ID_0 \times PT_0$.
Then $p = E[c_1]/(E[c_1] + E[c_0]) = ID_1 \times PT_1/(ID_1 \times PT_1 + ID_0 \times PT_0)$.
Divide above and below by $ID_0 \times PT_1$ to get $p = IDR/(IDR + PT_0/PT_1)$.
Invert this equation, i.e., express IDR as a function of p & PT_0/PT_1, as above.

[5] the reviewer is fictitious, but this type of fussy situation is not that unusual – the next question deals with an important rct of vaccination efficacy with even more extreme data.

v. Compare these exact CI's (based on binomial split of the total number of cases) with the log-based ones arrived at via the procedure in pp 137-8 of Rothman2002, and write a short response to the reviewer's concern. You might be tempted to say that statistical reviewers are overly precise about the wrong issues (as jh once said to the editor of the BMJ) but be more diplomatic that jh was!

4. When the log-based CI[6] won't even work – and we have to rely on *conditioning* to 'convert a 2-sample problem to a 1-random variable problem.'

Refer to the "**Efficacy Analyses of a Human Papillomavirus Type 16 L1 Virus-like-Particle Vaccine**" excerpt (example 17/31).

For the primary analysis of the data from this rct, the point estimate of the IDR was 0, i.e., the point estimate of the efficacy was 100%. Clearly IDR_L is also 0 (it can't be negative!), but what about IDR_U? If ones calculates the SE for a log-based CI using formula 7-5 (p 137), i.e., SE $=\sqrt{1/0+1/41}$, we get an error (or `Inf` if we type `sqrt(1/0+1/41)` into `R`!)

The same complication arises with a log-based CI for an OR from a 2×2 table with a zero entry. There, as here, one solution is to *condition* on the *total* of the two numerators, as we did with the numbers of accidents among pilots.

We have virtually an exact 50:50 split of the women-years in the vaccinated ($_1$) and placebo ($_0$) experience, i.e., $PT_0/PT_1 = 1076.9/1084.0 \approx 1$, so

$$IDR_U = \frac{p_U}{1-p_U} \times \frac{1076.9}{1084.0} \approx \frac{p_U}{1-p_U}.$$

i. Use the Excel spreadsheet – or interpolation in the table of exact binomial-based 95% CIs for a proportion – and the observation that $\hat{p} = 0/41$ of the cases arose from the index experience, to obtain an exact p_U.[7] From this, obtain IDR_U, and from this obtain the lower bound of the 95% CI for percent efficacy.

ii. From your experience thus far, and the size of the numerators involved, predict how close the log-based and (conditional) binomial-based CIs will be for the 'secondary efficacy analysis'? Compute both and test your prediction. [to get p_L and p_U, based on the 6/74, both Wilson and interpolation (between 6/70 and 6/75) within the table, should give close to exact binomial limits]

***History Corner (and soapbox!)*— Those who evaluate vaccines, e.g. in the 1954 salk Polio trial, don't report IDR, but rather (%) Efficacy = (100) × (1 - IDR).**

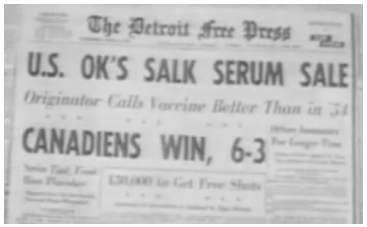

Figure 1: April 1955. from `http://www.polio.umich.edu/history/video.html`

5. Sloppy journalists plague our newspapers?

> It's a fact Quebecers are terrible drivers. According to statistics from the Société de l'assurance automobile du Québec, the province's automobile insurance board, Quebec has Canada's second-highest rate of death on the roads - 1,212 fatalities per 100,000 licensed drivers, compared with the Canadian average of 960. ["Forgetful drivers plague our roads", by Jason Magder, Montreal Gazette, Jan 12, 2008, p. A4]

(i) Correct this journalist.[8] (ii) Suggest another denominator than licensed-driver-years.

[6] described in Rothman2002 p 138, and in lines [13] and [17] in the second column of page 2 of jh handout "Statistical models for inference re epidemiologic parameters."

[7] If you try it, you will see that the upper limit of the Wilson CI is sensible, and pretty close to the upper limit of the exact one! Rothman2002 in exercise 2, p142 says the same – "it gives a *meaningful* upper limit". The exact upper limit of a $100(1-\alpha)\%$ binomial-based 2-sided CI for p, based on $0/n$, is $1-(\alpha/2)^{1/n}$. As Hanley and Lippman-Hand described, for $\alpha = 0.10$ and 0.05 respectively, this upper limit is close to $3/n$, and $3.7/n$ once $n > 20$

[8] You might wish to consult
`http://www.safety-council.org/info/traffic/stats2001.html` ,
`http://www.tc.gc.ca/roadsafety/stats/overview/2004/menu.htm` ,
`http://www.canadiandriver.com/news/041102-1.htm` .