Statistical Inference for Sex Ratio at Birth

1 Some Consideration for the Inference of SRB

High sex ratio at birth is one of hottest research topic in Asian countries such as in South Korea, India, and China etc. However, most studies use point estimates to estimate the population sex ratio at birth, rather than interval estimates, by using sample data. It is obviously inappropriate. For instance, based on 10% birth sample, a county enumerated 106 boys and 112 girls born in 2000 in China's population census; the sex ratio at birth (SRB) was 94.6. However, we found in the same county that in the third census, there were 2695 boys and 2377 girls born in 1981, and the SRB was 113.4; in the fourth census, there were 3074 boys and 2698 girls born in 1989, and the SRB was 113.9. Compared with the third and fourth censuses, it seems that the SRB in 2000 was not consistent. This is not true because the result of the 2000 census may be caused by random error, rather than the reality.

Some studies made inference for SRB, but the explanation of the methods in detail were rarely given (Hull 1990; Zeng 1993). Recently, we found that one calculates the confidence intervals for male proportion and for female proportion, respectively, and put the two interval estimates into the formula: SRB = male proportional interval/(1- male proportion interval). Obviously, such calculation is incorrect.

2. Estimates of the Standard Error and Confidence Interval

We know that the variance of new-born boys is

$$VAR_{Pm} = \frac{P_m(1 - P_m)}{n}$$

SRB is Odds of the male proportion (assume the number of female baby is one), that is,

$$SRB = \frac{P_m}{1 - P_m}$$

The variance of SRB is:

$$VAR_{SRB} = \left(\frac{\partial(SRB)}{\partial P_m}\right)^2 \times VAR_{Pm}$$

The standard error of SRB can be:

$$SE_{SRB} = \frac{1}{1 - P_m}\sqrt{\frac{SRB}{n}}$$

Which can be simplified to

$$SE_{SRB} = (1 + SRB) \cdot \sqrt{\frac{SRB}{n}}$$

The 95% confidence interval should be:

$$SRB \pm 1.96 \times SE_{SRB}$$

For example, we have

Case	SRB	N	SE	95%CI	
				Lower	Upper
1	104.9	1000	6.64	91.9	117.9
2	104.9	5000	2.97	99.1	110.7
3	119.3	1000	7.57	104.5	134.1
4	119.3	5000	3.39	112.7	125.9

3. One-Sample Test

[Question] We try to compare sample SRB with a given SRB in order to test if the SRB in a population is higher or lower than a given value.

We generally admit that the normal SRB should be in the interval 103-107, with middle value 105. We may test 1) the SRB for population is higher than 105; 2) the SRB for population is higher than the normal interval, that is, 107, testing if there is a male selection; and 3) the SRB for population is lower than 103, testing if there is a female selection. The formula is:

$$t = \frac{SRB_{sample} - SRB_{given}}{SE_{SRB}}$$

Suppose the sample SRB for an area is 115.5, the sample size of the birth is 2000. We try to test if the SRB is higher than the normal interval, that is, 107. We know that $SE_{SRB} = 5.18$, then

$$t = \frac{SRB_{sample} - SRB_{given}}{SE_{SRB}} = \frac{115.5 - 107}{5.18} = \frac{8.5}{5.18} = 1.64$$

As using one-tail test, we got the p value = 0.10, which is not significant.

If we want to make it significant, we have to increase the sample size. Suppose we expend the sample size to 3000. The $SE_{SRB} = 4.23$; t = 2.01 and p =0.044, which is significant.

4 Two-Sample Test

[Question] We will test if SRBs for two populations are significantly different through comparing two sample SRBs.

The are two methods for doing this, that is, 1) using Pearson Chi-square to test difference of two binary-category samples; 2) test null hypothesis：odds ratio = 1.

Example: Compare A sample with B, C, and D samples

	Variables and Statistic	A sample	Compared Areas：		
			B sample	C sample	D sample
	Male birth	3469	3184	3073	4853
Sample	Female birth	3250	2738	2698	4481
	SRB	106.74	116.29	113.90	108.30

First Method	$\chi^2(1)$	-	5.76	3.26	0.21
	p-value	-	0.016	0.071	0.650
Second Method	Odds Ratio	-	0.918	0.937	0.986
	95%CI	-	0.855	0.873	0.925
		-	0.985	1.006	1.050
	$\chi^2(1)$	-	5.76	3.26	0.21
	p-value	-	0.016	0.071	0.650

The two methods are consistent.

If the difference of SRB for two areas is affected by other variables, Mantel-Haenszel test should be used.

5. The Selection of the Sample Size

We only consider the type I error.

[Test 1] Test if the population SRB is normal when the sample SRB is normal. For instance, when the sample SRB = 105 and 95%CI, the probability for SRB>107 is:

N	2000	10000	30000
95%CI	96-114	101-109	103-107
Prob. For SRB>107	33%	16%	5%

[Test 2] Test if the population SRB is higher when the sample SRB is relatively higher. For instance, when the difference between sample SRB and 107 is statistically significant under 5% level, the sample size at least we need should be:

SRB	130	122	117	115	114	113	111
N	500	1000	2000	3000	4000	5000	10000

From the two tests above, we can see that testing for both the normal and abnormal SRBs is quite difficult and need large sample size.

The conclusion is that it is hard to successfully test the difference of SRBs even though the sample size is relatively large.