Erik Parner
3 April 2015

Solution to exercise 4-3

Background

It is a common belief that if the first two children in a family are of the same sex then the next child will most likely be of the same sex. All though scientific results are not completely consistent[1], most studies favor independence between the sex of the children in a family. Here we revisit the hypothesis using data from a random sample of the first and second born in 1000 Danish families.

Statistical methods

The chance of getting a boy the second time is analyzed using a binomial model and estimates with 95% confidence intervals (CI) are presented.

Results

1. Consider the chance of getting a boy the second time. Estimate this if the older sib was a boy and if she was a girl.

The changes of getting a boy the second time is 52.7% (95% CI: 48.2%-57.2%) if the first child is a girl and 46.5% (42.1%-51.0%) if the first child is a boy.

2. Estimate the risk difference comparing the chance of a boy the second time if the older sib was a boy to if the older sib was a girl. Test the hypothesis of no association between the sex of the first and the second child. Write a short conclusion including a discussion of the validity of the assumptions behind the calculations.

The chance of getting a boy the second time is 6.2% (0.0%-12.4%) higher if the first child is a girl as compared to when the first child is a boy. The data suggest that the chance of getting a boy the second time may depend on the sex of the first born child (p=0.050).

3. Estimate the proportion of boys among the oldest and youngest sibs.

The changes of getting a boy the first time is 50.3% (47.2%-53.4%) and the second time 49.6% (46.5%-52.7%).

4. Estimate the difference in the proportion of boys among the first and second child. Test the hypothesis of no difference. Write a short conclusion including a discussion of the validity of the assumptions behind the calculations.

The chance of getting a boy the second time is 0.7% (-3.9%-5.3%) lower as compared to chance of getting a boy the first time. The chance of getting a boy the first and second time is not statistical significant different (p=0.79).

5. Discuss the similarities and differences between what you looked at in 2 and 4.

The analysis performed in question 2 examines the hypothesis of independence between the sex of the first and second child, whereas the analysis performed in question 4 examines the hypothesis of the same chance of having the a boy the first and second time.

Conclusion

The chance of getting a boy the second time is 6.2% (12.4%-0.0%) higher if the first child is a girl as compared to when the first child is a boy. Although the data suggest that the chance of getting a boy the second time may depend on the sex of the first born child (p=0.050), the finding is in opposite direction as the hypothesis of same sex in the family.

References

1. http://en.wikipedia.org/wiki/Human_sex_ratio#Natural_factors

Do file

```
********************************************************************
* Solution to Exercise 4-3.
********************************************************************

cd "D:\Teaching\BasicBiostat\Exercises"

capture log close
log using "solution4-3.log",text replace

use siblings.dta, clear
codebook sex*
tabu sex1 sex2

* Many Stata commands expect 0/1 variables.
generate boy1=2-sex1 if sex1<.
generate boy2=2-sex2 if sex2<.
label define NoYes 0 "No" 1 "Yes"
label val boy1 NoYes
label val boy2 NoYes
tabu boy1 sex1
tabu boy2 sex2

* Q1.
* We use the tabulate command to get familiar with the counts,
* and the ci command to computed the chance of getting a boy the
* second time with 95% CI. The sex of the second child (boy2) is
* the outcome and the sex of the first child (boy1) is the exposure.
tabu boy1 boy2,row
ci prop boy2 if boy1==0
ci prop boy2 if boy1==1

* Q2
* Risk difference, and hypothesis of independence.
* Note:
* - Since the sex of the second child is the outcome we place
*   boy2 first in the cs command and then the exposure boy1.
* - The        cs command does not use the labels coded in boy1 and boy2,
*   but rather the general terminologi from cohort studies;
*   exposed (boy1=1, i.e. first born boy)
*   unexposed (boy1=0, i.e. first born girl)
*   cases (boy2=1, i.e. second born boy)
*   noncases (boy2=0, i.e. second born girl).
* - The command prtest and tabu also performs the independence test
*   as in the cs command.
cs boy2 boy1
prtest boy2 , by(boy1)
tabu boy1 boy2 , row chi2

* Q3
```

```
* Note: here the sex of the first (boy1) and second child (boy2) is
* the outcome.
ci prop boy1
ci prop boy2

* Q4
* We can't at this point assume that the sex of the
* two siblings are independent, so we should consider the
* data as paired data.
* Note:
* - The order of variable in mcc arbitrary, however we need in the
*   output to keep track of who is places where in the output table.
*   Here we use the 2. variable first and then the 1. variable.
* - The          mcc command does not use the labels coded in boy1 and boy2,
*   but rather the general terminologi from matched case-control
*   studies;
*   Cases (outcome of the second child, i.e. boy2)
*     exposed (boy2=1, i.e. second born boy)
*     unexposed (boy2=0, i.e. second born girl)
*   Control (outcome of the firt child, i.e. boy1)
*     cases (boy1=1, i.e. frist born boy)
*     noncases (boy1=0, i.e. first born girl).
* - Be careful: prtest makes a unpaired comparison, i.e. prtest boy1=boy2
mcc boy2 boy1


* Q5
* Same probability of the sex is not the same as independence.

log close
```