Chinese Fertility, Sex-ratios, and the 1-child Policy

Chinese Fertility, Sex-Ratios and the 1-child policy

In this lab we will look at results from the 1990 census. A 1% sample of the census has been made available by the IPUMS International project at the University of Minnesota. We've created a 10% sample of this 1% sample (a 1 in 1,000 sample) and prepared it for you. The data set is quite large, covering about 200,000 households.

We saw that, at the aggregate, the 1-child policy did not cause major changes in the level of fertility. However, we may be able to detect more subtle changes. Chinese parents may have reacted to the 1-CP not by reducing the number of children but by trying to influence the sex of their children either by sex-selective abortion, hiding girls when the census take came, or trying for another child when they only had girls (taking advantage of exceptions to the 1-CP).

In this lab we look at the sex-ratio at birth and how parents decisions to have more children may have depended on the sex composition of the children they already have. Using the 1990 census, we will also try to see if Chinese couples became more strategic regarding the sex of their children after the introduction of the 1-child policy in 1979.

Our goals in this analysis are

1. To look for signs of sex-preference based on the birth order and sex composition of the children that have been born
2. To see if the 1-child policy may have amplified son-preference.

We will look at the following indicators:

- Sex ratios of surviving children by year of birth to see if births became more masculine after the 1-child policy.
- Sex ratios by birth order (and year of birth) to see if higher order births are more likely to be boys.
- "Parity progression" by sex composition of children already born.

(Parity is the number of births a woman has had. So a woman of parity 3 has had 3 births. The parity progression ratio is defined as PPR_i, probability of having a child of order $i + 1$ among women of parity i. So, if $PPR_3 = .2$ it means that 20 percent of women who have 3 children "progress" to having a 4th birth.)

Part 0. Load the data

```
# leave this chunk as-is, but do press the green button to load quiz info
tot = 0
answer.key = "eJytVctu2zAQvOcrFulBCeDKtWMUbYEiyAtoLkXRFOgxoKWVREQiFZKK4n59Z0nZceAeegiQyCYp7s7OzK7zRzN0a
library(quizify)
source.coded.txt(answer.key)
```

```
df <- read.csv("/data175/china_1990_first_three_kids_by_age_and_sex.csv",
               na.string = "")
```

```r
## display
head(df)
```

```
##   hhserial age1 age2 age3 sex1 sex2 sex3    educmom    educpop    agric nkids
## 1    19000   17   NA   NA    f <NA> <NA> university university nonagric     1
## 2    26000    5   NA   NA    f <NA> <NA>  secondary university nonagric     1
## 3    50000   19   16   NA    m    f <NA>  secondary university nonagric     2
## 4    52000    2   NA   NA    m <NA> <NA> university university nonagric     1
## 5    70000   18   NA   NA    m <NA> <NA>    primary  secondary nonagric     1
## 6    71000    7   NA   NA    f <NA> <NA>  secondary  secondary nonagric     1
```

```r
## for convenience we assign the contents of this data.frame to
## individual vectors.
sex1 <- df$sex1
sex2 <- df$sex2
sex3 <- df$sex3
age1 <- df$age1
age2 <- df$age2
age3 <- df$age3
nkids <- df$nkids
```

- "hhserial" is the household number in the census. This can be ignored
- nkids is the number of children (defined as those under age 20)
- age1 is the age in years of the 1st born child (the oldest) in 1990
- age2 is the age in years of the 2nd born child in 1990
- age3 is the age in years of the 3rd born child in 1990
- sex1 is the sex of the 1st born child ("m" = "male"; "f" = "female")
- sex2 is the sex of the 2nd born child ("m" = "male"; "f" = "female")
- sex3 is the sex of the 3rd born child ("m" = "male"; "f" = "female")
- educmom is the educational attainment of the mother
- educpop is the educational attainment of the father
- agric is the "agricultural household type" (nonagric, agric)

  Note: We ignore educmom, educpop, and agric for this lab.

An "NA" means "not applicable" or "not available". Here an "NA" in age2 means there was only 1 child, and an "NA" in age3, following a number for age2 means there were only 2 children. (The same applies to NAs in sex2 and sex3).

We have listed only the first 3 births (omitting all higher order births). We have also dropped all of the households with zero children.

Some additional considerations are

> Young children (particularly those younger than school age) are thought to be underreported in the Chinese census. This underreporting is thought in part to be due to trying to avoid the fines associated with the 1-child policy.
>
> We are looking at the sexes of surviving children. We don't see those who were born and then who died.

Q0.1 Household number 50000 has

A. 19 boys and 16 girls

B. A 19 year old boy and a 16 year old girl ever born.

C. A 19 year old girl and a 16 year old boy ever born.

D. A 19 year old boy and a 16 year old girl reported in the census as living in their household.

```r
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer0.1 = "D"
quiz.check(answer0.1)
```

```
## Your  answer0.1 : D
## Correct.
## Explanation:  Both 'C' and 'D' are consistent with the data. But because these are the results of a l
```

```r
## Write your own code
## df[df$hhserial == somenumber,]
```

Part 1. The proportion female (by age and birth order)

We begin by looking at the proportion female π_f by age in 1990 and by birth order.

Our hypotheses is that the 1-child policy might have caused the proportion of female births to decline. It may have also encouraged underreporitng of surviving girls. Differential mortality and infanticide are not thought to be major causes of sex differences in surviving children.

Let's first look at the proportion female π_f by age in 1990

```r
get.prop.female <- function(sex)
{
    ## this function takes a vector of sexes ("m" and "f")
    ## and computes the proportion female
    round(prop.table(table(sex[!is.na(sex)]))["f"],4)
    ## Technical note: this is the original version of this function, which
    ## should now work fine since all the empty strings "" in the
    ## 'sex' variable have been read-in as "NA".
}

## usage example:

pi.1 <- get.prop.female(sex = sex1)     # returns proportion oldest
print(pi.1)                             # children (1st births that
```

```
##      f
## 0.4751
```

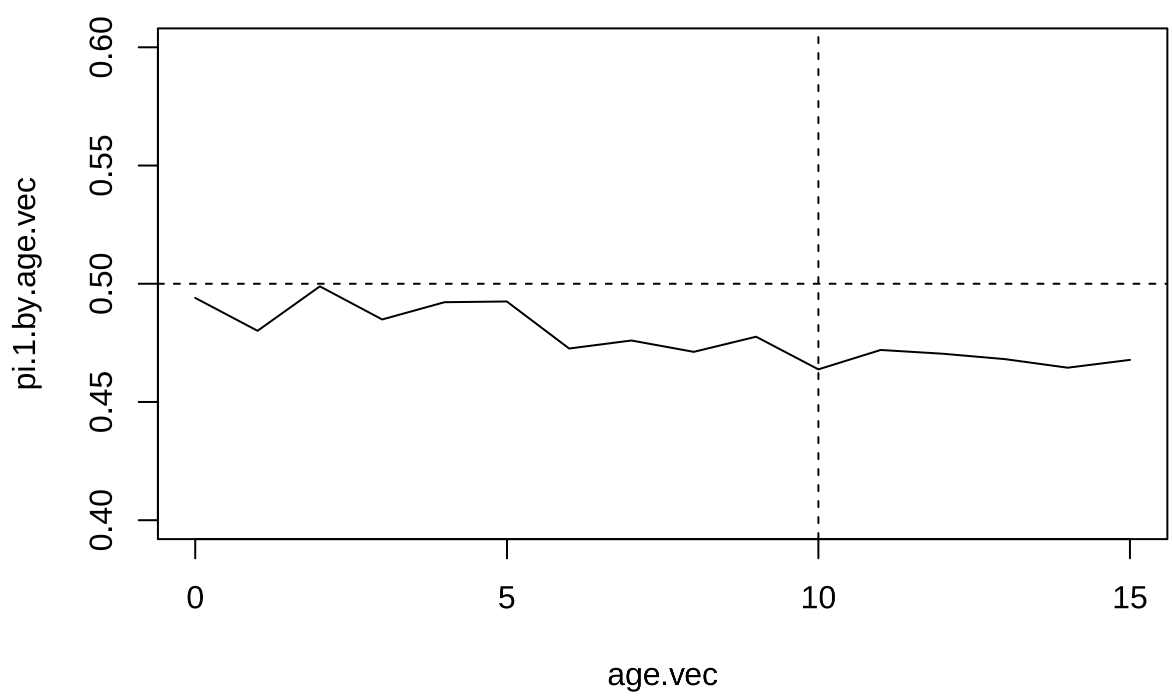
```r
                                        # are girls


age.vec = 0:15                          # this is a vector of ages in
                                        # 1990 that we are interested
                                        # in. We loop through these
                                        # values.

pi.1.by.age.vec <- NULL             # initialize a vector for storing
                                        # proportion female (pi) of
                                        # 1st births by age of oldest
                                        # child (1st birth)

## now we loop through the ages and get proportion female by age in 1990

for (i in 1:length(age.vec))
```

```
{
    this.age = age.vec[i]
    pi.1.by.age.vec[i] <- get.prop.female(sex1[age1 == this.age])
}

print(pi.1.by.age.vec)
```

```
##  [1] 0.4940 0.4801 0.4989 0.4849 0.4922 0.4925 0.4726 0.4760 0.4712 0.4776
## [11] 0.4638 0.4720 0.4704 0.4681 0.4645 0.4678
```

And plot results

```
plot(age.vec, pi.1.by.age.vec,
     type = "l",
     ylim = c(.4, .6))
abline(v = 10, lty = 2) # approximate begining of 1-CP (1990 - 10 =
                        # 1980 = about 9 months after 1-CP introduction
                        # in 1979
abline(h = .5, lty = 2)
```



> Interesting. It looks like the proportion of 1st births that are girls is increasing as we go forward in time (The figure shows decreases with age in 1990, but the older children were born earlier.) This is the opposite of what we expected.

Q1.1 We were expecting …

A. π_f to increase with age, because fewer girls would be born (or reported) after the 1-CP.

B. π_f to decrease with age, because fewer girls would be born (or reported) after the 1-CP.

```
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer0.2 = "B"
quiz.check(answer0.2)
```

```
## Your  answer0.2 : B
## Correct.
## Explanation:  'B' is correct because fewer girls (relative to boys) would
## .reduce pi_f.
```

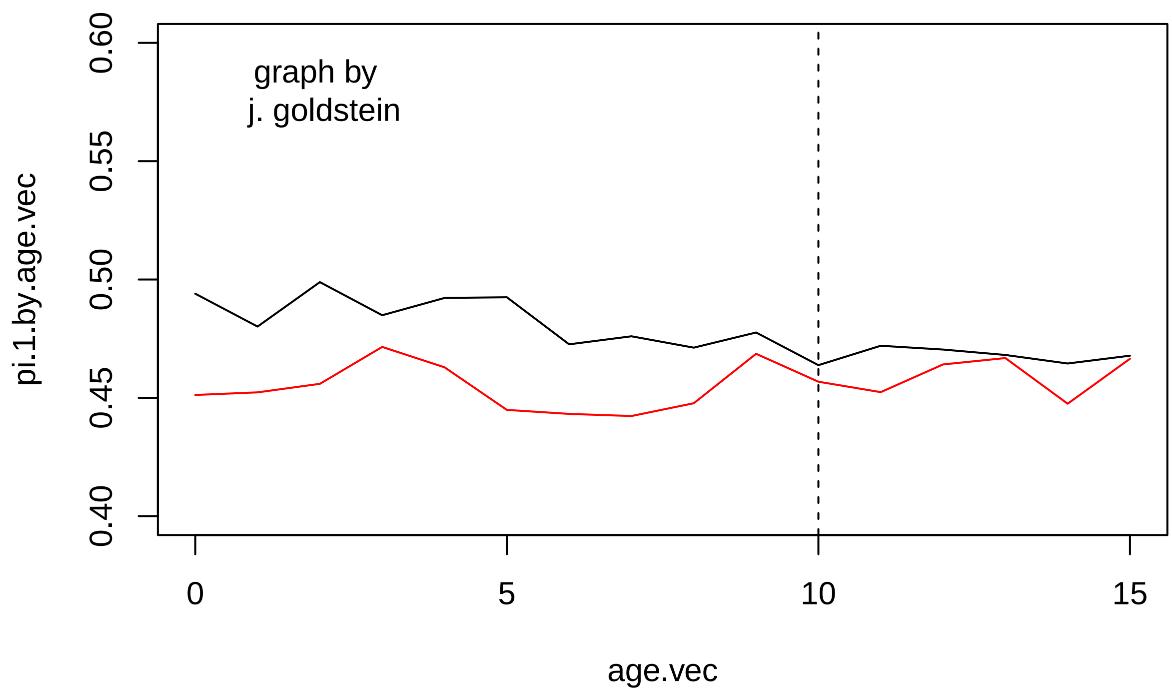
Let's see if the same suprising trend holds true for 2nd births.

We have copied and pasted the relevant code above to calculate "pi.2.by.age.vec" and plot this on the same graph using the following commands. Insert your name using the text function in the upper left. For the graded questions you will submit a copy of this graph.

```
pi.2.by.age.vec <- NULL             # initialize a vector for storing
                                        # proportion female (pi) of
                                        # 1st births by age of oldest
                                        # child (1st birth)

## now we loop through the ages and get proportion female by age in 1990
for (i in 1:length(age.vec))
{
    this.age = age.vec[i]
    pi.2.by.age.vec[i] <- get.prop.female(sex2[age2 == this.age])
}
print(pi.2.by.age.vec)
```

```
##  [1] 0.4512 0.4523 0.4559 0.4715 0.4629 0.4449 0.4432 0.4423 0.4477 0.4686
## [11] 0.4568 0.4524 0.4641 0.4668 0.4475 0.4665
```

```
plot(age.vec, pi.1.by.age.vec,
     type = "l",
     ylim = c(.4, .6))
lines(age.vec, pi.2.by.age.vec,
      type = "l", col = "red")
abline(v = 10, lty = 2) # approximate begining of 1-CP
text(2, .58, "graph by \n j. goldstein")
```




```
## text(x = 5, y = .42, "2nd births", col = "red")
## text(x = 5, y = .5, "1st births", col = "black")
## abline(h = .45, lty = 3, col = "red")
```

> Wow. Now we see that 2nd births are less likely than 1st births to be girls (and more likely to be boys). This effect is strongest for the younger children, born after the 1-CP.

Q1.2 The 1-CP appears to have …

A. Reduced the proportion female of 1st births.

B. Reduced the proportion female of 2nd births.

C. Reduced the proportion female of 2nd children reported in the census

D. Done nothing. There is no trend in π_2 by age.

E. It's complicated. In order to see something, we need to consider the difference in proportion female between 1st and 2nd births, before and after the 1 child policy.

```
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer0.3 = "E"
quiz.check(answer0.3)
```

```
## Your  answer0.3 : E
## Correct.
## Explanation:  'E' We can see there is a growing gap between the proportion female in 1st and 2nd bir
```

Sex ratio by birth order (and age)

To finish up this part of the analysis let's look to see

1. If the proportion female declines with birth order (without accounting for age)
2. If the proportion female declines with birth order and with the 1-CP (accounting for age)

Without age:

```
print("sex1")
```

```
## [1] "sex1"
```

```
pi.1 = get.prop.female(df$sex1)
print(pi.1)
```

```
##      f
## 0.4751
```

```
print("sex2")
```

```
## [1] "sex2"
```

```
pi.2 = get.prop.female(df$sex2)
print(pi.2)
```

```
##      f
## 0.4567
```

```
print("sex3")
```

```
## [1] "sex3"
```

```
## insert command for pi.3 = in next line

# print(pi.3) # un-comment this line to print
```

Q1.3 What happens to the proportion female by birth order.

A. It falls consistently, indicating more sex-selection in favor of sons for higher-order births.

B. It falls consistently, indicating more sex-selection (or under-reporting) in favor of sons for higher-order births.

C. There is no clear pattern

D. It falls from 1st to 2nd births, but rises for 3rd births.

```
##  "Replace the NA with your answer (e.g., 'A' in quotes)"
answer0.4 = "B"
quiz.check(answer0.4)
```

```
## Your  answer0.4 : B
## Correct.
## Explanation:  'B' is correct. Because this is a census, the sex ratio could also be influenced by rep
```

With age:

Here we cut-and-paste the code above to create "pi.2.by.age.vec" and "pi.3.by.age.vec"

```
## insert code here for creating pi.2.by.age.vec
```

```
## insert code here for plotting pi.1.age.vec, pi.2.by.age.vec, and
## pi.3.age.vec all on same graph (with pi.3.age.vec in "blue")
```

Q1.4 What does adding 3rd births do to the story?

A. Not much. The pattern is too noisy.

B. It re-inforces what we've already found, that there is sex-selection for higher births that might have increased after 1-CP

C. It contradicts what we've already found.

For discussion (No auto-correct answers available.)

Part 2. A closer look. How the decision to have another child depends on sex of children already born.

The evidence we have looked at above is fairly indirect, averaging across all families. We saw that the higher order births were more likely to be boys than girls.

We can go further and look at variation between households. We can take advantage of the fact that the first child is sometimes a boy and sometimes a girl and see if those who have a girl are more likely to "try again" for a boy.

Parity progression to 2, based on sex of birth 1.

We now calculate $PPR1$ (the progression from parity 1 to a second birth). We do this twice, once for those whose first child was a boy and once for those whose first child was a girl. We include age of the 1st child so that we take into account that it takes time to have a conceive and give birth to a 2nd child.

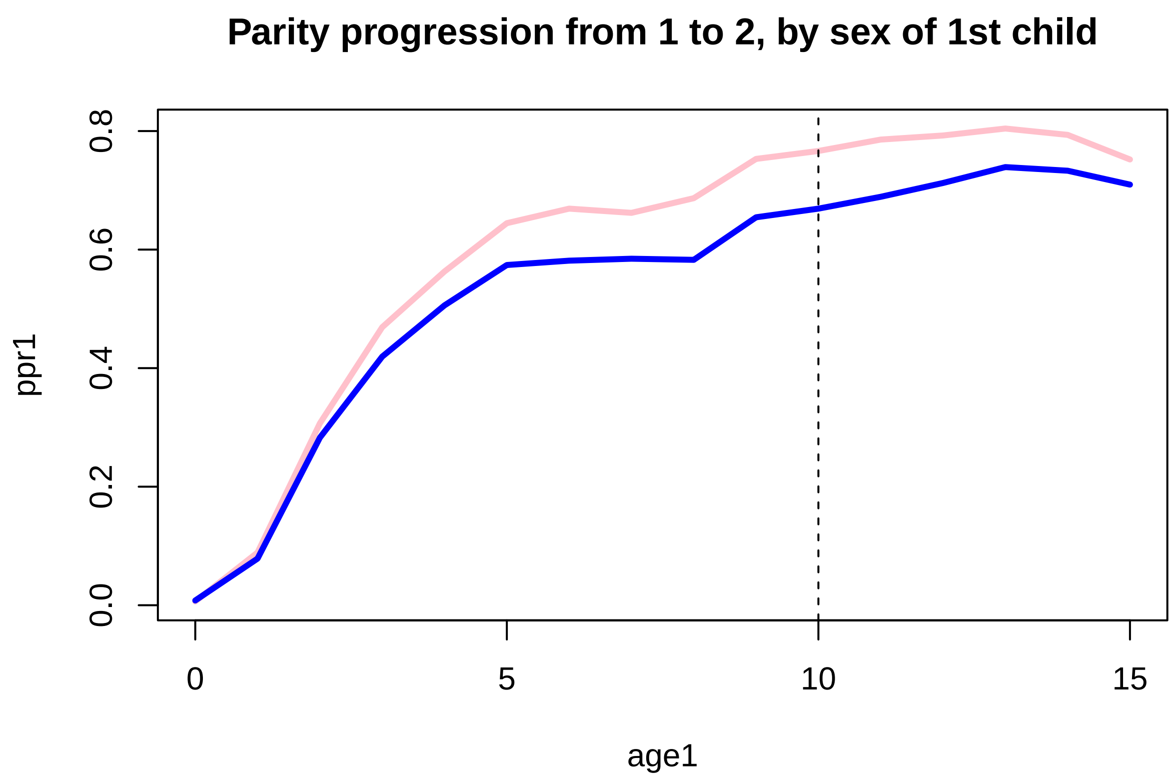
We show how to do this in full for 2nd births. For your assignment you will look at 3rd births.

```
get.ppr.by.parity <- function(nkids, parity)
{
    num = sum(nkids > parity)          # number of hh in nkids vec that
                                        # have another birth

    denom = length(nkids)               # number of hh in nkids vector
    ppr = num/denom
    return(ppr)
}


ppr1.sex1f.vec = NULL                # ppr for families with sex1 = "f"
ppr1.sex1m.vec = NULL                # ppr for families with sex1 = "m"

for (i in 1:length(age.vec))
{
    this.age = age.vec[i]
    ppr1.sex1f.vec[i] <- get.ppr.by.parity(nkids[sex1 == "f" &
                                                 age1 == this.age],
```

```
                                              parity = 1)
    ppr1.sex1m.vec[i] <- get.ppr.by.parity(nkids[sex1 == "m" &
                                                    age1 == this.age],
                                              parity = 1)
}
```

Now let's plot

```
plot(age.vec, ppr1.sex1f.vec,
     col = "pink", lwd = 3, type = "l",
     ylab = "ppr1",
     xlab = "age1",
     main = "Parity progression from 1 to 2, by sex of 1st child")
lines(age.vec, ppr1.sex1m.vec, col = "blue", lwd = 3, type = "l")
abline(v = 10, lty = 2)
```



Q2.1 Does this make sense?

A. Yes, it means that those with a female first born are more likely to try again.

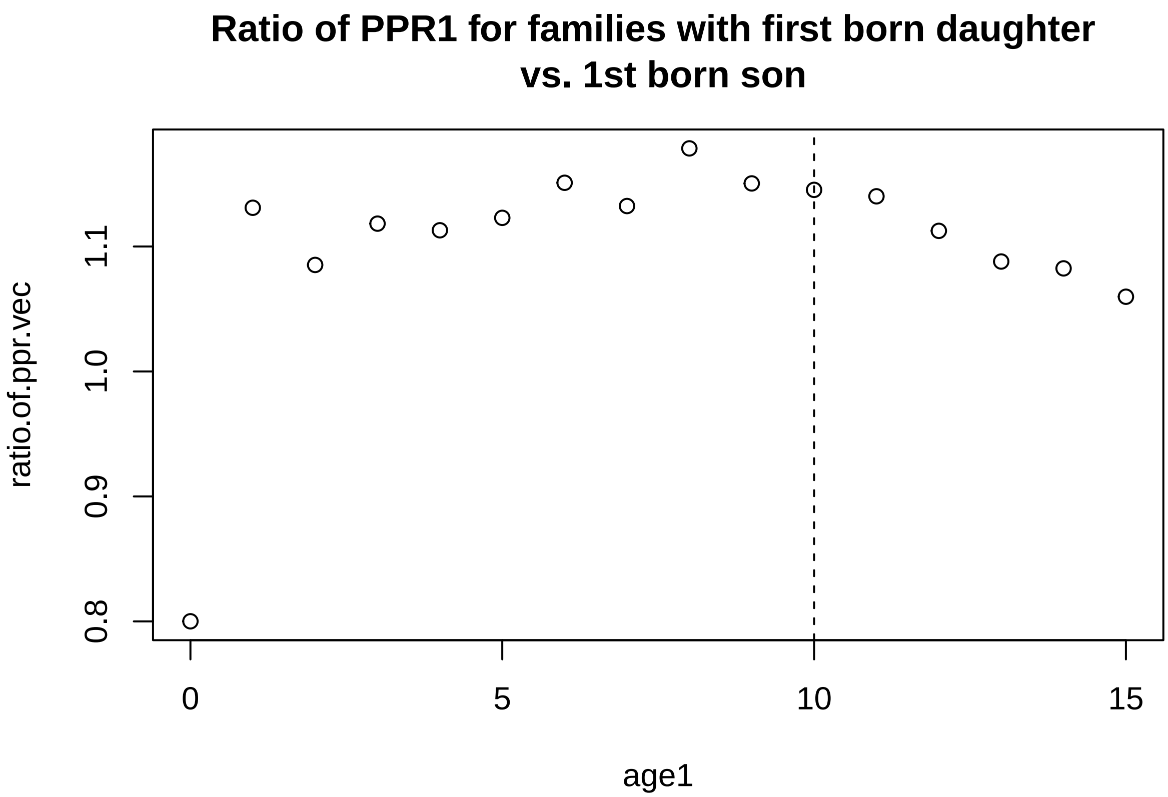
B. Yes, it means that those with a female first born are less likely to try again.

C. No, the gap shouldn't disappear for the younger children, since that is after the 1 child policy

For discussion (No auto-correct answers available.)

The age of the first child is influenced both by "period effects" from the year of birth and by "age effects" on the time it takes to have another child. We can remove the age effect somewhat by looking at the ratio of ppr by sex of the first born.

```r
ratio.of.ppr.vec = ppr1.sex1f.vec/ppr1.sex1m.vec
plot(age.vec, ratio.of.ppr.vec,
     xlab = "age1",
     main = "Ratio of PPR1 for families with first born daughter \n vs. 1st born son")
abline(v = 10, lty = 2)
```



Q2.2 What conclusion do we draw?

A. We should ignore the data point for age1 equals zero, since there has been too little time to have a 2nd birth.

B. Having a girl 1st encourages, on average, Chinese parents to try again.

C. The effects of son preference are largely stable over time

D. There is a slight suggestion that that the effect of son preference is stronger after the 1-CP

E. All of the above.

No auto-correct answer. Discuss in class.

Congratulations. We have finished the in-class lab.

Graded Question

Extend the analysis in Part 2 to look at the parity progression ratios from 2nd to 3rd births depending on the sex of the first two children.

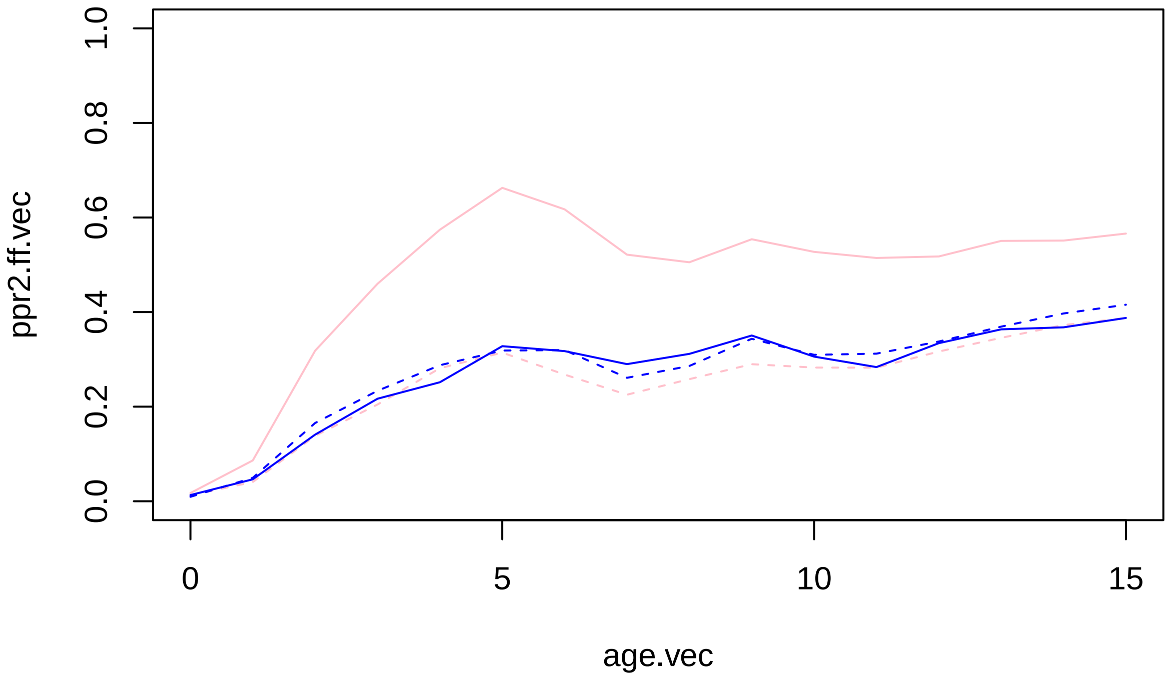
1. Create vectors

ppr2.ff.vec ppr2.fm.vec ppr2.mf.vec ppr2.mm.vec

2. Loop through ages of the 2nd child and calculate ppr2 values, filling up these vectors.

Here is some code for calculating ppr2.ff.vec and ppr2.fm.vec. We've marked where you need to insert additional code in order to create the remaining two vectors. (Hint: This is not meant to be very hard. You just need to copy the right code and make the right changes.)

```
ppr2.ff.vec <- NULL
ppr2.fm.vec <- NULL
ppr2.mf.vec <- NULL
ppr2.mm.vec <- NULL
## insert additional code here
for (i in 1:length(age.vec))
{
    this.age = age.vec[i]
    ppr2.ff.vec[i] <- get.ppr.by.parity(nkids[sex1 == "f" &
                                              sex2 == "f" &
                                              !is.na(sex2) & # see note1
                                              age2 == this.age],
                                        parity = 2)
    ppr2.fm.vec[i] <- get.ppr.by.parity(nkids[sex1 == "f" &
                                              sex2 == "m" &
                                              !is.na(sex2) & # see note1
                                              age2 == this.age],
                                        parity = 2)
    ppr2.mf.vec[i] <- get.ppr.by.parity(nkids[sex1 == "m" &
                                              sex2 == "f" &
                                              !is.na(sex2) & # see note1
                                              age2 == this.age],
                                        parity = 2)
    ppr2.mm.vec[i] <- get.ppr.by.parity(nkids[sex1 == "m" &
                                              sex2 == "m" &
                                              !is.na(sex2) & # see note1
                                              age2 == this.age],
                                        parity = 2)
    ## insert additional code here
}

plot(age.vec, ppr2.ff.vec, col = "pink", lty = 1, type = "l",
     ylim = c(0, 1))
lines(age.vec, ppr2.fm.vec, col = "pink", lty = 2)

lines(age.vec, ppr2.mf.vec, col ="blue", lty = 2)
lines(age.vec, ppr2.mm.vec, col ="blue", lty = 1)
```




```
## insert additional code here
```

(Note1: for R-gurus only, the line `!is.na(sex2)` means "where sex2 is not equal to NA". Since sex2 is NA when only 1 child has been born, this is equivalent to saying "consider only the households with at least two children".)

3. Plot your results, so that all of the vectors are on the same graph. Use the following plotting options

ppr2.ff.vec: ' col ="pink", lty = 1 ' (solid pink) ppr2.fm.vec: ' col ="pink", lty = 2 ' (dashed pink) ppr2.mf.vec: ' col ="blue", lty = 2 ' (dashed blue) ppr2.mm.vec: ' col ="blue", lty = 1 ' (solid blue)

Put your name on the plot (anywhere that looks fine) and turn in a copy of the plot.

4. Which of the following appears to be the tendency among Chinese parents who have already had two children? A. They prefer at least 1 boy and 1 girl. B. They prefer at least 1 boy. C. They prefer at least 1 girl. [Choose the best answer and explain < 50 words]

They prefer at least 1 boy. We see that parents with two females have higher parity progression ratio. It seems to imply that they are willing ot have a third child to get a boy. After all, it is the male who carries out the family name in Chinese culture.

5. Does the analysis of 3rd births reinforce or contradict our analysis of 2nd births? [< 50 words]

No, not really. It actually reinforces it. The parity progression ratio is higher in general if the birth is given to a female. This leads to another attempt for a male, but then a female. Then another attempt for a male that leads to a 3rd birth.

Congratulations. You've finished the graded portion of the lab on China.