

## Cohort fertility

R code for today's class can be obtained as <http://minato.sip21c.org/demography-special/code20140508.R>.

Cohort fertility can be calculated from the same data as several period fertility measures, using "synthetic cohort".

Parity Progression Ratio (PPR) is also important concept. Birth interval analysis is more difficult, but can effectively use a small sample data.

### Cohort and period fertility analysis

- Cohort analysis considers the experience of one group of people over time (usually all those born or marrying during a particular time interval; the former is called as "birth cohort", the latter is called as "marriage cohort").
- Three big problems: (1) data must be in a form of long, consistent time series, and such data are rare (2) censoring problem, the later experience of young or more recent cohorts is unknown, (3) human thinking is much more familiar with cross-sectional way than cohort way.
- See Table 5.1 (ASFRs in England and Wales, 1940-1985). Each column shows period ASFRs for the specific year. When comparing horizontally, changes of ASFRs for each age-group can be seen. The women aged 15-19 in 1940 become 20-24 in 1945, so that the horizontal comparison is done between different population. **By considering diagonals of the table, we can compare the fertility of different birth cohorts.** (cf. Actually each ASFR is affected by all of age, period, and birth cohort, so that APC (age-period-cohort) models are sometimes applied.)
- **Cohort fertility rate (CFR)** for a birth cohort = the sum of ASFRs which the birth cohort experiences divided by 1,000 and multiplied by 5 (age-intervals of each ASFRs). It means the average number of children which women in the particular birth cohort will have borne by age 50. For the 1920-1924 birth cohort in Table 5.1,

```
(15+103+136+84+46+13+1)/1000*5 # CFR for 1920-4 birth cohort in England and Wales
T5.1 <- read.delim("http://minato.sip21c.org/demography-special/EnglandWales1940.txt")
colSums(T5.1[,-1])*5/1000 # TFR
sum(diag(as.matrix(T5.1[,2:8])))*5/1000 # CFR for 1920-4 birth cohort
# This calculation assumes that the ASFRs in 1940 were experienced throughout 1940-1944, which may cause errors.
sum(diag(as.matrix(T5.1[,3:9])))*5/1000 # CFR for 1925-9 birth cohort
sum(diag(as.matrix(T5.1[,4:10])))*5/1000 # CFR for 1930-4 birth cohort
sum(diag(as.matrix(T5.1[,5:11])))*5/1000 # CFR for 1935-9 birth cohort
sum(diag(as.matrix(T5.1[,6:11])))*5/1000 # CFR for 1940-44 birth cohort, 45-49 is censored.
```

- For Japanese data, since ASFRs are given for each age, it must be summed up for 5 year age-groups.
  - **library(fmsb)**  
`x <- function(i) {tapply(Jfert[,i],(Jfert$Age-10)%/%5,sum)}`  
`y <- Vectorize(x)`  
`z <- y(2:14)`  
`sum(diag(z[, 1:8])) # CFR for 1935-1939 birth cohort of Japanese females`  
`sum(diag(z[, 5:12])) # CFR for 1955-1959 birth cohort of Japanese females`
- Sometimes CFR is called as 'Completed Family Size', but it's misleading. Based on the maternal recall data, we can calculate the average number of children born per woman aged 50-54, which is different from CFR, because only the woman who survived to 50-54 can answer.
- Comparing with TFR, CFR is generally less variable (see, Table 5.1 and 5.2), because TFR is affected by tendency of couples to change the timing of their births, CFR is not (If couples have their children earlier, TFR will rise, but CFR will not change). The effect of change of timing is also known as "**tempo effect**". When the change in TFR is observed, it will be the big problem whether it is truly the change in fertility level or only the change of timing. The answer is, it cannot be known until the relevant cohort complete their childbearing.
  - As an extreme example of "tempo effect", Japanese "*Hinoe-Uma* (year of the white horse)" (丙午) **superstition** (A girl born in that year is believed to be going to grow up ill-tempered and no one would want to marry her, because she will kill and eat her husband) in 1966 is well known. The TFRs changed as 2.2, 1.6, and 2.1 during 1965-1967.
  - As another example, in Romania in 1966, sudden introduction of a more restrictive abortion law made birth rate arisen, and the adoption of contraception made it fallen again.
  - As yet another example, in Britain during 1946-1947, the post-war "baby-boom" is recorded due to the postponed births during the war.
- Such events may cause less effect on CFR. The influences on fertility which do affect cohort rates are called as "**quantum effect**". It inevitably affects the period fertility, so that the analysis become difficult. The change in CFR can be analyzed using life-table technique.

### The manipulation and presentation of cohort data

- There are various ways to show cohort fertility. Period fertility tables (Table 5.1) can be changed to show the cohort appear as rows (Table 5.3).
- The assumption in this table is that the period fertility observed in a particular year stays unchanged for the whole period of 5 years: For example, 1940 ASFR is applicable for the whole period from 1940 to 1944.
- If we use the data classified only by age, we cannot precisely translate ages to a year of birth: eg. A woman states 20 years old at some time in 1980, she may be born on a date between 1 Jan 1959 to 31 Dec 1960, ranges 2 years. In the opposite direction (trying to analyze TFR by date of birth), similar problem may occur.
- The process of family building can easily be seen by cumulating the figures of Table 5.3 over age like Table 5.4.

### Marriage cohorts

- The member in any birth cohort may marry at different ages
- Using marriage cohort (women marrying in the same year, irrespective with age), the effect of the differences of the age of marriage can be adjusted.
- The MDSFR (marriage-duration specific fertility rate) can be defined below.  

$$\text{MDSFR} = \frac{\text{Births in year } x \text{ to women who married in year } y}{\text{Total women who married in year } y}$$
- When we cumulate MDSFR over the marriage duration, we can get cumulated fertility by marriage cohort like Table 5.5.
- Comparing with Table 5.4, Table 5.5 shows that: down the side, date of marriage, along the top, marriage duration.
- Problems: (1) difficulty to treat divorce and remarriage (sometimes only dealing with unbroken first marriage may solve the problem, but may fail in the society where the divorces and remarriages are common), (2) Within a single marriage cohort, the age of marriage may largely vary. A woman marrying young tends to have higher fertility than one marrying late, (3) Most of births occur in the first 1-2 year after marriage for first births, but longer for higher-order births. To solve it, the additional disaggregation of the marriage cohort specific fertility table like Table 5.5 by birth-order is used.

### Synthetic cohorts

- TFR is sometimes interpreted as "the number of children a woman would have if she survived to age 50 and experienced over her reproductive years exactly the given ASFRs" (Newell C: "Methods and models in demography" The Guilford Press, 1988, p.57), but no real group actually experience any particular set of ASFRs.
- A hypothetical group who are supposed to experience a set of ASFRs is "synthetic cohort".
- Using this idea, TFR is imaginary experience of a synthetic cohort, CFR is actual experience of a birth (or marriage) cohort.
- Lifetable analysis also applies synthetic cohort.

### Parity Progression Ratios

- Besides vital data, the census or survey-derived data can also contribute to fertility analysis as Child/Woman Ratio (CWR), and Parity Progression Ratio (PPR).
- **PPR is the probability of having another child given that one has already had a certain number.**
- (eg.) If a woman has already had three births, PPR is the probability of her having a fourth.
- Example calculation is shown in Table 5.6: The first column is the number of children ever born. The second column is the number of corresponding women aged 45-49 (who are considered as completed fertility). Then, the figures in the second column are cumulated from bottom to give the number of women with at least  $n$  children ever borne (CEB) to give the third column. Lastly, we divide adjacent figures in the third columns to give PPRs ( $a_0$  means the proportion of women in the cohort who become mothers,  $a_1$  means the proportion of women having second baby among the all mothers, and so on). CFR is an arithmetic series of products of PPRs.  

$$\text{CFR} = a_0 + a_0a_1 + a_0a_1a_2 + \dots$$

### Birth intervals

- PPR only considers numbers of children. Timing of having babies has much more information of fertility.
- Fertility may rise if the age at first birth become earlier or the birth intervals between successive births become shorter.
- The definition of birth interval: the time between successive live births and usually measured in months. Stillbirths/abortions are not counted. Different from inter-pregnancy intervals (the period from the end of one pregnancy to the start of the next, which is used to measure the length of exposure to the risk of conception).
- The first birth interval = the time between the first and second live births. Not the interval from marriage to first birth
- Post-partum amenorrhoea, prior to the return of menstruation after confinement, lasts at least 1.5 months and can be extended by prolonged lactation.
- The technique of survival analysis can be applicable for dealing with censored data like **Kaplan-Meyer method**. Maternity history chart can be drawn using **median birth intervals** based on birth intervals data.

- **library(fmsb)**  
**example(mhchart)**