Original article

Statistical Analysis of Female Proportion among the Japanese Infant Population of 0–4 Year Olds

Teruyuki Nakayama*1, Kenichi Satoh*2, and Megu Ohtaki*2

*1Department of Mathematics, Faculty of Science, Hiroshima University, Higashi-Hiroshima 739–8526, Japan
*2Department of Environmetrics and Biometrics, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1–2–3, Kasumi, Minami-ku, Hiroshima 734–8553, Japan

We deal with the problem of whether or not there is an increase in female proportion among 0–4 year olds in the Japanese population during the last 21 years. The validity of the use of a binomial model for analyzing the municipality-specific data of female proportion is examined using a beta-binomial model. It is suggested from a logistic regression analysis that there exists a difference of the female proportion of infant population among population-size-specific groups of municipalities. The highest proportion was observed for the rural group, and the lowest one was observed for the city group. A slight but statistically significant increasing time-trend in female proportion among the city group is detected via a test with the null hypothesis that the proportions of females are all equal against an ordered alternative.

Key words: Beta-binomial distribution; Binomial distribution; Isotonic test; National census; Ordered alternative; Sex ratio; Test of equality.

1. Introduction

It has been worried that the male proportion among newborn babies might be decreasing as a result of pollution in the environment. Davis et al. (1998) reported that the ratio of male to female births was decreasing in several industrial countries: Denmark (1950–1994), Netherlands (1950–1994), Canada (1970–1990) and the United States (1970–1990). We consider here, using a test of equality, the problem whether or not there is an increasing time-trend in female proportion among 0–4 year olds in the Japanese population, based on the data collected by the national census from 1975 to 1995 every five years. A scatter plot of the proportion of females 0–4 years of age in 1995 with its mean and the 95% confidence bounds based on a homogeneous binomial model is illustrated in Figure 1, and similar plots can be obtained from the data of the other years. It shows that the proportions of females for small populations are fairly disperse, but the others can be assumed to have small dispersions. Therefore, considering the het-
T. Nakayama, K. Satoh, and M. Ohtaki

Fig. 1. The proportion of females 0–4 year olds in the Japanese population in 1995.

Table 1. The number of municipalities included in each group divided on the basis of total population in 1995.

<table>
<thead>
<tr>
<th>Year</th>
<th>[1,300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>469</td>
<td>1,442</td>
<td>1,458</td>
<td>3,369</td>
</tr>
<tr>
<td>1980</td>
<td>573</td>
<td>1,419</td>
<td>1,397</td>
<td>3,389</td>
</tr>
<tr>
<td>1985</td>
<td>666</td>
<td>1,429</td>
<td>1,283</td>
<td>3,378</td>
</tr>
<tr>
<td>1990</td>
<td>856</td>
<td>1,372</td>
<td>1,148</td>
<td>3,376</td>
</tr>
<tr>
<td>1995</td>
<td>1,065</td>
<td>1,233</td>
<td>1,072</td>
<td>3,370</td>
</tr>
</tbody>
</table>

It is assumed that the number of females, $X_i$, are independent random variables, each of
Table 2. The number of females and population size for 0–4 year olds in Japan.

<table>
<thead>
<tr>
<th>Year</th>
<th>[1, 300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>42,407</td>
<td>417,826</td>
<td>4,435,794</td>
<td>4,896,027</td>
</tr>
<tr>
<td>1980</td>
<td>86,689</td>
<td>855,104</td>
<td>9,105,633</td>
<td>10,047,426</td>
</tr>
<tr>
<td>1985</td>
<td>50,958</td>
<td>410,361</td>
<td>3,741,880</td>
<td>4,203,199</td>
</tr>
<tr>
<td>1990</td>
<td>104,307</td>
<td>842,008</td>
<td>7,679,768</td>
<td>8,626,083</td>
</tr>
<tr>
<td>1995</td>
<td>57,944</td>
<td>406,121</td>
<td>3,176,421</td>
<td>3,640,486</td>
</tr>
<tr>
<td></td>
<td>118,656</td>
<td>831,352</td>
<td>6,509,255</td>
<td>7,459,263</td>
</tr>
<tr>
<td></td>
<td>72,183</td>
<td>372,952</td>
<td>2,721,723</td>
<td>3,166,858</td>
</tr>
<tr>
<td></td>
<td>147,826</td>
<td>764,337</td>
<td>5,580,734</td>
<td>6,492,897</td>
</tr>
<tr>
<td></td>
<td>86,321</td>
<td>330,578</td>
<td>2,508,340</td>
<td>2,925,239</td>
</tr>
<tr>
<td></td>
<td>176,325</td>
<td>676,424</td>
<td>5,142,505</td>
<td>5,995,254</td>
</tr>
</tbody>
</table>

which follows a binomial distribution with true proportion of females $p_i$ and population size $n_i$ ($i=1, \cdots, m$), where $m$ is the number of municipalities. The probability density function (p.d.f.) of $X_i$ can be expressed as

$$f_{g}(x|i,p) = \frac{n_i \cdot C_x \cdot p^x \cdot (1-p)^{n_i-x}}{\mathcal{B}(a,b)}$$  \hspace{1cm} (1)

where $n_i \cdot C_x = n_i! \cdot (n_i-1) \cdots (n_i-x_i+1)!$. The maximum likelihood estimator (MLE) of $p_i$ is given by the proportion of females in the $i$-th city, $\hat{p}_i = X_i/n_i$. Under the binomial distribution, the null hypothesis of equal proportions can be written as

$$H_0: p_1 = \cdots = p_m$$  \hspace{1cm} (2)

In Section 2, we consider a beta-binomial distribution for the alternative hypothesis to examine the validity of use of the binomial model. The homogeneity in the proportion of females among the three groups [1, 300), [300, 1,000) and [1,000, ∞) is also tested. In Section 3, we analyze the data using a logistic regression model. Finally, the increasing time-trend in the proportion of females at 0–4 year of age is examined using a test of equality against an ordered alternative in Section 4.

2. Validity of the binomial model

In this section, the problem of whether the use of the binomial model is valid or not under the dispersion in the proportions of females such as that in Figure 1, is examined. To set up the alternative hypothesis against the null hypothesis (2), we consider a beta-binomial distribution. It is assumed that the true proportions of females, $p_i$, are independent and identical random variables, each of which follows a beta distribution with two unknown parameters $a, b > 0$. The p.d.f. is expressed as

$$g_{\beta}(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{\mathcal{B}(a,b)}$$  \hspace{1cm} (3)
where \(0<\rho<1\), \(\mathcal{B}(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)\) and \(\Gamma(\alpha)=\int_0^\infty t^{\alpha-1}e^{-t}dt\). The mean and the variance are

\[
E_\beta(p)=\mu_\beta=\frac{a}{a+b}, \quad \text{Var}_\beta(p)=\sigma_\beta^2=\frac{ab}{(a+b)^2(a+b+1)}.
\] (4)

The variables \(\mu_\beta\) and \(\sigma_\beta^2\) have a one-to-one relation with the two unknown parameters \(a\) and \(b\), as follows.

\[
a=-\mu_\beta+\frac{\mu_\beta^2(1-\mu_\beta)}{\sigma_\beta^2}, \quad b=-\mu_\beta-1+\frac{\mu_\beta(1-\mu_\beta)^2}{\sigma_\beta^2}.
\] (5)

From (1) and (3), it can be assumed that the number of females, \(X_i\), are independent random variables, each of which follows a beta-binomial distribution. The p.d.f. can be expressed as

\[
f_{\beta\beta}(x_i\mid a,b)=E_\beta\left\{f_\beta(x_i\mid p_i)\right\} = \int_0^1 f_\beta(x_i\mid p_i)\theta_\beta(p_i\mid a,b)dp_i
\]

\[
= n C_{x_i} \frac{\mathcal{B}(x_i+a,n_i-x_i+b)}{\mathcal{B}(a,b)}.
\] (6)

The mean and the variance are

\[
E_{\beta\beta}(X_i)=\frac{n_i a}{a+b}, \quad \text{Var}_{\beta\beta}(X_i)=\frac{n_i a b (a+b+n_i)}{(a+b)^2(a+b+1)}.
\]

From (5), we can express (6) in terms of \(\mu_\beta\) and \(\sigma_\beta^2\) as

\[
f_{\beta\beta}(x_i\mid \mu_\beta, \sigma_\beta^2)=n C_{x_i} \frac{\mathcal{B}\left(x_i-\mu_\beta+\frac{\mu_\beta^2(1-\mu_\beta)}{\sigma_\beta^2}, n_i-x_i+\mu_\beta-1+\frac{\mu_\beta(1-\mu_\beta)^2}{\sigma_\beta^2}\right)}{\mathcal{B}\left(-\mu_\beta+\frac{\mu_\beta^2(1-\mu_\beta)}{\sigma_\beta^2}, \mu_\beta-1+\frac{\mu_\beta(1-\mu_\beta)^2}{\sigma_\beta^2}\right)}.
\]

Note that \(f_{\beta\beta}(\ast\mid \mu_\beta, 0)\) is equivalent to \(f_{\beta}(\ast\mid \mu_\beta)\).

The MLE’s of \(a\) and \(b\) under the model (6) are obtained by maximizing the logarithm of the likelihood function

\[
l_{\beta\beta}(a,b)=\sum_{i=1}^n \log f_{\beta\beta}(x_i\mid a,b).
\] (7)

As the MLE’s of \(a\) and \(b\) are given, the MLE’s of \(\mu_\beta\) and \(\sigma_\beta^2\) can also be obtained through (4). Let the MLE’s of \(\mu_\beta\) and \(\sigma_\beta^2\) be \(\mu\) and \(\sigma^2\), respectively.
Here, we examine the validity of the use of binomial model. The hypothesis is expressed as

\[ H_0 : p_1 = \cdots = p_m \quad \text{v.s.} \quad K_0 : p_1, \cdots, p_m - i.i.d. \quad \beta(\mu_\beta, \sigma^2_\beta), \quad \sigma^2_\beta > 0, \]  

(8)

where \( \beta(\mu_\beta, \sigma^2_\beta) \) denotes a beta distribution with mean \( \mu_\beta \) and variance \( \sigma^2_\beta \). This hypothesis (8) is equivalent to the hypothesis,

\[ H_0 : X_i \sim B(n_i, p_0) \quad \text{v.s.} \quad K_0 : X_i \sim \beta B(m_\beta, \sigma^2_\beta), \quad \sigma^2_\beta > 0, \]  

(9)

where \( p_0 = p_1 = \cdots = p_m, B(n_i, p_0) \) and \( \beta B(\mu_\beta, \sigma^2_\beta) \) denote a binomial and a beta-binomial distribution, respectively. The null hypothesis \( H_0 \) in (9) is equivalent to \( X_i \sim \beta B(p_0, 0) \) for any \( i \). Therefore, under the hypothesis (9), the log likelihood ratio statistic is defined as

\[
T = -2l_{\beta B}(\hat{p}_0, 0) + 2l_{\beta B}(\mu_\beta, \sigma^2_\beta),
\]  

(10)

where \( l_{\beta B}(p_0, 0) = \sum_{i=1}^m \log f_B(x_i | p_0) \), \( l_{\beta B}(\mu_\beta, \sigma^2_\beta) = \sum_{i=1}^m \log f_{\beta B}(x_i | \mu_\beta, \sigma^2_\beta) \), and

\[
\hat{p}_0 = \frac{1}{N} \sum_{i=1}^m X_i, \quad N = \sum_{i=1}^m n_i.
\]  

(11)

Since \( \mu_\beta \) is equal to \( p_0 \) in the case of \( \sigma^2_\beta = 0 \), under the beta-binomial model, the hypothesis (9) is also expressed as

\[ H_0 : \sigma^2_\beta = 0 \quad \text{v.s.} \quad K_0 : \sigma^2_\beta > 0. \]  

(12)

We apply \( \hat{p}_0 \) as an estimator not only for \( p_0 \) but \( \mu_\beta \). Here we use the more simple test statistic \( T^* \) instead of \( T \) in (10), which is defined as

\[
T^* = -2l_{\beta B}(\hat{p}_0, 0) + 2l_{\beta B}(\hat{p}_0, \hat{\sigma}^2_\beta),
\]  

(13)

Since the parameter value under the null hypothesis \( H_0 \) is on the boundary of the parameter space, the asymptotic distribution of the statistic \( T^* \) is not that of \( X_1^2 \). However, we can obtain the asymptotic distribution as in Self and Liang (1987, Case 5). Its p.d.f. is

\[
g(t) = \begin{cases} 
0, & \text{if } t < 0 \\
0.5\{\delta(t) + f(t)\}, & \text{if } t \geq 0
\end{cases}
\]

where \( \delta(t) \) denotes the Dirac delta function and \( f(t) \) is the p.d.f. of the \( X_1^2 \)-distribution

\[
f(t) = (2\pi)^{-1/2}e^{-t^2/2}, \quad t \geq 0.
\]

The cumulative distribution function can be written as

\[
F(T^* \leq t) = \begin{cases} 
0, & \text{if } t < 0 \\
0.5 + 0.5\int_{0}^{t} f(x)dx, & \text{if } t \geq 0
\end{cases}
\]  

(14)
Table 3. The proportion of females in each group.

<table>
<thead>
<tr>
<th>Year</th>
<th>[1, 300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.4892</td>
<td>0.4886</td>
<td>0.4871</td>
<td>0.4873</td>
</tr>
<tr>
<td>1980</td>
<td>0.4885</td>
<td>0.4874</td>
<td>0.4872</td>
<td>0.4873</td>
</tr>
<tr>
<td>1985</td>
<td>0.4883</td>
<td>0.4885</td>
<td>0.4880</td>
<td>0.4880</td>
</tr>
<tr>
<td>1990</td>
<td>0.4883</td>
<td>0.4879</td>
<td>0.4877</td>
<td>0.4877</td>
</tr>
<tr>
<td>1995</td>
<td>0.4896</td>
<td>0.4887</td>
<td>0.4878</td>
<td>0.4879</td>
</tr>
</tbody>
</table>

Table 4. MLEs of the standard deviations in the beta distribution.

<table>
<thead>
<tr>
<th>Year</th>
<th>[1, 300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.00639</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00026</td>
</tr>
<tr>
<td>1980</td>
<td>0.00000</td>
<td>0.00575</td>
<td>0.0018</td>
<td>0.00044</td>
</tr>
<tr>
<td>1985</td>
<td>0.00000</td>
<td>0.00410</td>
<td>0.00130</td>
<td>0.00134</td>
</tr>
<tr>
<td>1990</td>
<td>0.00753</td>
<td>0.00000</td>
<td>0.00133</td>
<td>0.00126</td>
</tr>
<tr>
<td>1995</td>
<td>0.00652</td>
<td>0.00480</td>
<td>0.00056</td>
<td>0.00078</td>
</tr>
</tbody>
</table>

Table 5. The statistics $T^*$ for a test of validity with respect to the use of binomial model.

<table>
<thead>
<tr>
<th>Year</th>
<th>[1, 300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.24088</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00743</td>
</tr>
<tr>
<td>1980</td>
<td>0.00000</td>
<td>4.39936</td>
<td>0.00088</td>
<td>0.03230</td>
</tr>
<tr>
<td>1985</td>
<td>0.00000</td>
<td>1.20447</td>
<td>1.59165</td>
<td>1.78388</td>
</tr>
<tr>
<td>1990</td>
<td>0.76465</td>
<td>0.00000</td>
<td>1.20486</td>
<td>0.99275</td>
</tr>
<tr>
<td>1995</td>
<td>0.52676</td>
<td>1.68935</td>
<td>0.04035</td>
<td>0.15131</td>
</tr>
</tbody>
</table>

The proportions of females among 0–4 year olds, are shown in Table 3. These values are equivalent to the MLE's of $p_0$ under $H_0$, which are given by (11). The values of $\hat{\mu}_\beta$ and $\hat{\sigma}_\beta^2$ are taken by maximizing (7) and using (4). Note that we apply $\hat{p}_0$ in Table 3 as $\hat{\mu}_\beta$. The standard deviations $\sigma_\beta$ used in (13) are shown in Table 4. The test statistics $T^*$ for the hypothesis (12) are shown in Table 5.

Each $p$-value corresponding to the statistics $T^*$ obtained using (14) is shown in Table 6. It is easily seen that the null hypothesis $H_0$ for the group $[300, 1,000)$ in 1980 is rejected. However, since almost all the $p$-values are more than 0.05, we accept the null hypothesis $H_0$. Therefore, it is concluded that the use of the binomial model is valid for these data.

The test problem (12) is considered by dividing the data into three groups $[1,300)$, $[300, 1,000)$ and $[1,000, \infty)$ for each investigated year. Now, we examine homogeneity on the proportions of females for these groups and years. From the above test, it can be assumed that
Table 6. $p$-values corresponding to the statistics $T^*$. 

<table>
<thead>
<tr>
<th>Year</th>
<th>[1, 300)</th>
<th>[300, 1,000)</th>
<th>[1,000, ∞)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>0.312</td>
<td>0.500</td>
<td>0.500</td>
<td>0.466</td>
</tr>
<tr>
<td>1980</td>
<td>0.500</td>
<td>0.018</td>
<td>0.488</td>
<td>0.429</td>
</tr>
<tr>
<td>1985</td>
<td>0.500</td>
<td>0.136</td>
<td>0.104</td>
<td>0.091</td>
</tr>
<tr>
<td>1990</td>
<td>0.191</td>
<td>0.500</td>
<td>0.136</td>
<td>0.160</td>
</tr>
<tr>
<td>1995</td>
<td>0.234</td>
<td>0.097</td>
<td>0.420</td>
<td>0.349</td>
</tr>
</tbody>
</table>

the random variable $Y = \sum_{i=1}^{m} X_i$ follows a binomial distribution with parameters $p_0$ and $N = \sum_{i=1}^{m} n_i$.

3. Logistic regression analysis

Based on the census data of five times divided into three groups, let $Y_{jk}$ be a random variable for the $j$-th investigated year and the $k$-th group ($j=1, \cdots, 5$, $k=1, 2, 3$). It can be assumed that the $Y_{jk}$ are independent, each of which follows a binomial distribution with true proportion of females $q_{jk}$ and population size $N_{jk}$. Here, we apply a logistic regression model with quadratic time denoted as

$$
\log \frac{q_{jk}}{1-q_{jk}} = \beta_0 + \sum_{r=1}^{9} \beta_r z_r,
$$

where $\beta_0$ and $\beta_r$'s are unknown parameters, $z_1, z_2, z_3$ are dummy variables depending on the $k$-th group, respectively, i.e.,

$$
z_1 = \begin{cases} 1 & k=1 \\ 0 & \text{else} \end{cases}, \quad z_2 = \begin{cases} 1 & k=2 \\ 0 & \text{else} \end{cases}, \quad (z_3 = 1 - z_1 - z_2 \text{ is omitted in the analysis}),
$$

and $z_r (r=4, \cdots, 9)$ are variables depending on the $j$-th investigated year and the $k$-th group, respectively,

$$
z_4 = z_1 t_j, \quad z_5 = z_2 t_j, \quad z_6 = z_3 t_j, \quad z_7 = z_4 t_j^2, \quad z_8 = z_5 t_j^2, \quad z_9 = z_6 t_j^2,
$$

$$
t_j = \{-10, j=1; \ -5, j=2; \ 0, j=3; \ 5, j=4; \ 10, j=5\}.
$$

After fitting all combinations of polynomials included in model (15), we found the best model selected by AIC to be

$$
\log \frac{q_{jk}}{1-q_{jk}} = \hat{\beta}_0 + \hat{\beta}_1 z_1 + \hat{\beta}_2 z_2 + \hat{\beta}_3 z_6 + \hat{\beta}_4 z_9,
$$

where $\hat{\beta}_r$ is the estimator of $\beta_r$. The values of $\hat{\beta}_r$ and their corresponding 95% confidence intervals are shown in Table 7, and the fitted model (16) is illustrated in Figure 2. It can be seen that there exists a significant difference ($p<0.05$) in the female proportion of the infant population.
Table 7. The estimates of $\beta_i$ and their 95% confidence intervals for the fitted logistic model selected by AIC.

<table>
<thead>
<tr>
<th>Estimates</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>$-4.917 \times 10^{-2}$, $-4.816 \times 10^{-2}$</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>$4.422 \times 10^{-3}$, $9.193 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$2.020 \times 10^{-3}$, $4.139 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>$1.362 \times 10^{-4}$, $2.280 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>$-1.149 \times 10^{-5}$, $3.861 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

Fig. 2. The fitted model on the proportions of females for five investigated years and three groups.

among population-size-specific groups of municipalities $[1,300)$, $[300,1,000)$ and $[1000,\infty)$, where the highest proportion was observed for the rural group $[1,300)$, and the lowest one was observed for the city group $[1,000,\infty)$. There also exists a slight increasing trend in the female proportion of the infant population during the period from 1975 to 1994 in the city group.

4. Time-Trend Test

In this section, the problem whether there is an increasing time-trend on the proportion of
females among the 0–4 year olds population is examined. Since the null hypothesis $H_0$ given by
(12) is accepted in the preceding section, it can be assumed that each of the numbers of females, $X_i$, is independently distributed as a binomial with $(n_i, p_0)$ $(i = 1, \ldots, m)$.

Before testing, we overview the time-trend from estimated female proportions and their confidence intervals under the null hypothesis. The mean and the variance of $P_0$ given by (11) are $E_B(\hat{P}_0|H_0) = p_0$ and $\text{Var}_B(\hat{P}_0|H_0) = p_0(1 - p_0)/N$, respectively. According to the central limit theorem,
\[
\frac{\sqrt{N} (\hat{P}_0 - P_0)}{\sqrt{p_0(1 - p_0)}} \xrightarrow{d} N(0,1), \quad N \to +\infty.
\]

The 100(1-$\alpha$)% confidence interval of $p_0$ is asymptotically given by
\[
\left[ \hat{P}_0 - u_{a/2} \sqrt{\frac{\hat{P}_0(1 - \hat{P}_0)}{N}}, \hat{P}_0 - u_{a/2} \sqrt{\frac{\hat{P}_0(1 - \hat{P}_0)}{N}} \right], \quad (17)
\]
where $u_\alpha$ is the upper $\alpha$ point of the standard normal distribution.

Using the census data, the 95% confidence interval of $p_0$ can be obtained from (17), which for $\alpha = 0.05$ is based on $u_{0.025} = 1.96$. These values are shown in Table 8, in which the upper and lower rows for each year represent the upper and the lower intervals, respectively. Figure 3 shows $\hat{p}_0$ and its 95% confidence intervals. It shows that the confidence intervals for $p_0$ for [1, 300] are wider than the others.

Recall that $Y_{jk}$ is a random variable for the $j$-th investigated year and the $k$-th group. For each fixed $k$, it can be assumed that the $Y_j$’s are independent random variables, each of which follows a binomial with $q_j$ and $N_j(j = 1, \ldots, 5)$. Then we consider the following test problem:

$$H_2: q_1 = q_2 = \cdots = q_5,$$
Several tests for the hypothesis (18) are given by Kulatunga et al. (1996). The problem is examined using the method proposed by Bartholomew (1959) and Shorack (1967), which is referred to as an “isotonic test” by Collings et al. (1981). The statistic is given by

\[ T_{II} = \sum_{j=1}^{5} \frac{N_j \left( \hat{q}_j^* - \hat{q}_0 \right)^2}{\hat{q}_0 (1 - \hat{q}_0)} , \]

where \( \hat{q}_0 = \sum_{j=1}^{5} Y_j / \sum_{j=1}^{5} N_j \) and \( \hat{q}_j^* \) is the isotonic regression of \( \hat{q}_j \) with weights \( N_j \). If the sample sizes satisfy the condition that \( N_j/N^* \rightarrow w_j \in (0, 1) \) with \( N^* = \sum_{j=1}^{5} N_j \rightarrow \infty \), for \( j = 1, \cdots, 5 \), then the asymptotic null distribution of the above statistic is given as follows. If \( H_2 \) is true, then for \( c > 0 \)}
Table 9. The statistic $T_{II}$ and the $p$-value for a test of equality with respect to time-trend on the proportion of female.

<table>
<thead>
<tr>
<th></th>
<th>[1,300)</th>
<th>(300, 1,000)</th>
<th>[1,000, $\infty$)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{II}$</td>
<td>0.5358</td>
<td>0.9977</td>
<td>13.9805</td>
<td>15.4300</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.5013</td>
<td>0.3878</td>
<td>0.0006</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

$$\lim_{N* \to \infty} \Pr(T_{II} \geq c) = \sum_{r=2}^{5} P(\tau, 5; w) \Pr(\chi^2_{r-1} \geq c),$$  \hspace{1cm} (19)$$

where $P(\tau, 5; w)$'s are the level probabilities determined only by $w=(w_1, \ldots, w_3)$ (for detail, e.g., Robertson et al. (1988)) and $\chi^2_{r-1}$ is a chi-square random variable with $r-1$ degrees of freedom.

Using the census data, the isotonic regressions $\hat{q}_r^*$, the level probabilities $P(\tau, 5; w)$ and the value of $T_{II}$ are obtained. Its $p$-value can be easily obtained through (19). The values of $T_{II}$ and $p$-values for each group are shown in Table 9. It is easily seen that the null hypotheses $H_2$ overall and for the group [1,000, $\infty$) are rejected. Therefore, it is concluded that the proportions of females overall and in the group [1,000, $\infty$) are increasing.

5. Discussion

The primary purpose of our study is to examine whether or not there exists an increasing trend in the female proportion at birth recently in Japan. For this purpose, it would be better to analyze the sex-specific number of births rather than the sex-specific population of 0–4 year olds. The reason why we dealt with the data of 0–4 years is that no municipality-specific data on the number of sex-specific births are available. Therefore, it should be remarked that there exist several limitations to the interpretation of the results of this study.

Although it is difficult for us to give a proper explanation for the cause of the difference of female proportion among population-size specific group of municipalities, that is, the lower rates observed in the larger city, we can appeal to some background situations which may influence the municipality specific female proportion among the infant population. It is well known that the mortality among male infants is usually higher than that among females [e.g. Fukutomi (1989)]. In urban area, there are many hospitals and clinics which can provide modern medical treatment, so that a city dweller has a higher chance of getting high quality medical examination and treatment. It can be deduced that the mortality among male infants becomes lower in urban areas compared with rural areas, which might yield the decrease in female proportion among infant population in the city group. On the other hand, our analysis showed a finding contrary to our expectation in a sense, that is, there exists a slight increasing trend the female proportion among the infant population during the period from 1975 to 1994 in the city group. Although this trend may coincide with the recent declines in the male proportion at birth in sev-
eral industrial countries [Davis et al. (1998)], we cannot judge at the present whether or not there exists some relationship between an increase in the female proportion among the infant population and pollution in the environment. Further detailed investigation involving not only birth but also stillbirth statistics data is necessary for solving this problem.

Acknowledgements

The authors wish to thank Prof. Manabu Sato, Prof. Eiji Yamamoto, Prof. Takemi Yanagimoto, Prof. Takashi Yanagawa and Dr. John Cologne for their helpful comments. We also thank the referees for some very useful suggestions. This work was supported in part by Grants-in-Aid for Scientific Research provided by the Ministry of Welfare.

REFERENCES


