* **Brief introduction to research problem**

In developing countries, mortality of children is unacceptably high. Many studies have been conducted to investigate the relationship between short birth interval and child mortality over the years. Ties and George found that there is casual relationship between these two, to be more specific, short birth interval does associate with higher mortality risks. Further results from the paper also suggest that short birth spacing is related to the overlap of gestation and lactation, which causes great stress for the mother and the child; and women with short intervals are often not have chance to attend prenatal care. Sian at el 1993 goes further by examining the hierarchy structure of the data, investigating the family effects in relationship between short birth interval and child mortality. By building a random effects logistic model, the family effects are found to be highly significant, and the relationship between child mortality and birth spacing is still robust after controlling for the correlation between siblings.

The purpose of this research is to make an extension of the problem above: that is, to investigate whether a similar relationship exists between short birth interval and nutritional status as measured by ‘weight-for-age z-scores”. The z-score is the standard deviation above or below the mean. To the extent of DHS survey, z-score is calculated by standardising the sample data against the WHO standard population. Children who are very large or very small can’t be accurately tracked on the standard grow charts and therefore their weight needs to be converted to z-score. Hence, we will focus on ‘weight-for-age z-scores” as a measure of nutritious status and consider it as the dependent variable in our model.

* **Initial Analysis of the data**

***Waz*** (weight for age z-score) has a total of 4,083 values, with a minimum of -5.77 (the lightest child has weight being 5.77 standard deviation below the standard mean according to WHO) and a maximum of 5.70. Overall, children in the sample data tends to be underweight as *waz* has mean value of -1.2748. Similarly, mean value for child age ***(cage)*** is 17.11 (months), with the oldest child being 35 months old, and youngest just 1 month old. Standard deviation of child’s age is 9.936. Mean number of children under 5 to be 2.40 children under 5/per family with standard deviation of 1.738.

**Table

Description automatically generated**

Similarly, mean value for child age ***(cage)*** is 17.11 (months), with the oldest child being 35 months old, and youngest just 1 month old. Standard deviation of child’s age is 9.936. Mean number of children under 5 to be 2.40 children under 5/per family with standard deviation of 1.738.

|  |  |
| --- | --- |
| **Table  Description automatically generated** | **Table  Description automatically generated** |
| Table  Description automatically generated with low confidence |  |

Most of the baby size at birth ***(babysz)*** is considered “large/normal”, which account for 91.9% out of 4,083 children. 92.8% of baby are delivered on time, and only 1.8% are premature. Nearly half of children in the survey are delivered at home (50%) and half of them are delivered at hospital (49%). 848 children out of 4083 children are first birth in the family, account for 20.8% in total. 490 children were born within 24 months from the birth of the first child (32.8%), and 2,745 children were born out of 24 months interval from the previous child’s birth, accounting for 67.2%.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Size of baby at birth** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | large/normal | 3753 | 91.9 | 91.9 | 91.9 | | small | 330 | 8.1 | 8.1 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Diarrhoea in 2 weeks prior to survey** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | no | 3347 | 82.0 | 82.0 | 82.0 | | yes | 736 | 18.0 | 18.0 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Fever in 2 weeks prior to survey** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | no | 2536 | 62.1 | 62.1 | 62.1 | | yes | 1547 | 37.9 | 37.9 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Preceeding birth interval** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | First Birth | 848 | 20.8 | 20.8 | 20.8 | | < 24 months | 490 | 12.0 | 12.0 | 32.8 | | >= 24 months | 2745 | 67.2 | 67.2 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Mothers Education** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | Below Primary | 1422 | 34.8 | 34.8 | 34.8 | | Primary | 2524 | 61.8 | 61.8 | 96.6 | | Secondary (or above) | 137 | 3.4 | 3.4 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Mother's Occupation** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | None | 1138 | 27.9 | 27.9 | 27.9 | | Manual | 2848 | 69.8 | 69.8 | 97.6 | | Non-manual | 97 | 2.4 | 2.4 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Toilet Facilities** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | None | 774 | 19.0 | 19.0 | 19.0 | | Pit | 3256 | 79.7 | 79.7 | 98.7 | | Flush | 53 | 1.3 | 1.3 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Household Wealth** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | Low | 1956 | 47.9 | 47.9 | 47.9 | | Medium | 1133 | 27.7 | 27.7 | 75.7 | | High | 994 | 24.3 | 24.3 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Region** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | coastal | 877 | 21.5 | 21.5 | 21.5 | | northern highlands | 331 | 8.1 | 8.1 | 29.6 | | lake | 1519 | 37.2 | 37.2 | 66.8 | | central | 381 | 9.3 | 9.3 | 76.1 | | southern highlands | 553 | 13.5 | 13.5 | 89.7 | | south | 422 | 10.3 | 10.3 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Respondant's Religion** | | | | | | |  | | Frequency | Percent | Valid Percent | Cumulative Percent | | Valid | other | 22 | .5 | .5 | .5 | | moslem | 1268 | 31.1 | 31.1 | 31.6 | | catholic | 1246 | 30.5 | 30.5 | 62.1 | | protestant | 947 | 23.2 | 23.2 | 85.3 | | none | 600 | 14.7 | 14.7 | 100.0 | | Total | 4083 | 100.0 | 100.0 |  | |
|  |  |
|  |  |
|  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Size of baby at birth** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | large/normal | 3753 | 91.9 | 91.9 | 91.9 |
| small | 330 | 8.1 | 8.1 | 100.0 |
| Total | 4083 | 100.0 | 100.0 |  |

Using boxplot plotting against *waz* score, it can be seen that large/normal baby size at birth has a relatively higher weight-for-age mean score compared to baby who was born small.

There is no significant differences in mean score of *waz* between children of first birth, children born less than 24 months preceding birth interval and more than 24 months preceding birth interval.

37.9% of mothers has fever in 2 weeks prior to the survey, while the rate for diarrhoea is 18%. The majority of mother education level is primary school, which account for 61.8% of the total. There is 34.8% of women has education below primary school and only 3.4% of women has secondary (or above) school. Similarly, majority of partner’s education are primary education (59.4%). The percentage for partner’s education of below primary and secondary school is 23.2% and 7.0% respectively. 10% remaining in partner’s education category was explained by either partner’s missing education information or mother has no partner. In terms of occupation, 69.8% of mother’s occupation is manual, 27.9% is none and 2.4% is non-manual.

From the boxplot, *waz* score of children whose mother has no fever or diarrhoea 2 weeks prior to the survey is slightly higher than children whose mother has fever/ diarrhoea 2 weeks prior to the survey. Children whose mother having secondary tends to have higher waz score than those whose mother only has below primary education or primary education. Mothers who have non-manual occupations also tends to have higher waz score children.

Most of family household use pit toilet (79.7%), 1.3% of household use flush and 19% of total don’t have a toilet at home. Majority was defined as “low income” (47.9%). Medium and high-income household hold a relatively equal rate of 27.7% and 24.3% respectively. Majority of household lives near lake area (37.2%), followed by coastal (21.5%) and southern highlands (13.5%).

Boxplot showing that for children in household which has flush toilet tends to have average waz score higher than children in household using pit or have no toilet. Children in Northern Highlands area tends to have highest mean of waz score, followed by one living in Lake area and Southern Highland.

By fitting regression models using waz as the response variable, regress on the number of covariates, we are able to find estimate of regression coefficients and standard errors, as well as identify significant variables with the p-value < 0.05 in the model, which are: *Cage, babysz, diarr, fever, mother’s education, mother’s occupation, toilet, wealth, religion.*

However, adjusted R-square value of this model is only 11.35%. Residuals plot also indicates that violation of normality assumption and independent variances assumption.

The dataset contains 4,083 children who are clustered within 3,713 women, within 349 clusters. There are many possible reasons why weight-for-age score tends to cluster in families. Some household can have different diet for their children, and siblings tend to have the same home environment conditions, which consequently contribute to their weight grow. Suppose children with same mother are in the same household, we can investigate the effect of clustering in family on the relationship between birth interval and weight-for-age score.

|  |  |
| --- | --- |
| **Chart, scatter chart  Description automatically generated** | **Chart, box and whisker chart  Description automatically generated** |
| **Chart, box and whisker chart  Description automatically generated** | Chart, box and whisker chart  Description automatically generated |
| Chart, box and whisker chart  Description automatically generated | Chart, box and whisker chart  Description automatically generated |
| Chart, box and whisker chart  Description automatically generated |  |
|  |  |
|  |  |

The possible existence of a family effect can be investigated by producing a scatter plot of the mean of *waz* score for a child for each woman against women.

Chart, scatter chart

Description automatically generated

The scatter plot suggest that this mean varies considerably across mother. This suggest that the previous model ignore the existence of a potentially family effect, and it also underestimates the standard errors for the model parameters. As a result, fitting a multilevel model with two levels: *childid* and *momid* can be suitable in this case.