

# A framework of fetal age and weight estimation

Phuoc-Loc Nguyen<sup>1</sup>, Thu-Hang Ho-Thi<sup>2</sup>

<sup>1</sup>Huong Duong Company, Ho Chi Minh City, Vietnam

<sup>2</sup>Vinh Long General Hospital, Vinh Long Province, Vietnam

## Email address:

ng\_phloc@yahoo.com (Phuoc-Loc Nguyen), bshangvl2000@yahoo.com (Thu-Hang Ho-Thi)

## To cite this article:

Phuoc-Loc Nguyen, Thu-Hang Ho-Thi. A Framework of Fetal Age and Weight Estimation. *Journal of Gynecology and Obstetrics*. Vol. 2, No. 2, 2014, pp. 20-25. doi: 10.11648/j.jgo.20140202.13

---

**Abstract:** Fetal age and weight estimation plays the important role in pregnant treatments. There are many estimate formulas created by the combination of statistics and obstetrics. However, such formulas give optimal estimation if and only if they are applied into specified community or ethnic group with characteristics of such ethnic group. This paper proposes a framework that supports scientists to discover and create new formulas more appropriate to community or region where scientists do their research. The discovery algorithm used inside the framework is the core of the architecture of framework. This algorithm is based on heuristic assumptions, which aims to produce good estimate formula as fast as possible. Moreover, the framework gives facilities to scientists for exploiting useful information under pregnant statistical data.

**Keywords:** Fetal Age Estimation, Fetal Weight Estimation, Regression Model, Estimate Formula

---

## 1. Introduction

Fetal age and weight estimation is to predict the birth weight or birth age before delivery. It is very important for doctors to diagnose abnormal or diseased cases so that she/he can decide treatments on such cases. Because this paper mentions both age estimation and weight estimation, for convenience, the term “birth estimation” implicates both of them. There are two methods for fetal estimation:

- Calculating volume of fetal inside mother womb and basing on such volume and the mass density of flesh and bone, it is easy to calculate fetal weight.
- Applying statistical regression model: fetal ultrasound measures such as bi-parietal diameter (*bpd*), head circumference (*hc*), abdominal circumference (*ac*) and fetal length (*fl*) are recorded and considered as the input sample for regression analysis which results in a *regression function*. This function is the formula for estimating fetal age and weight according to ultrasound measures such as *bpd*, *hc*, *ac* and *fl*. Data is composed of these ultrasound measures is called gestational sample or statistical sample. Terms: “sample”, “data” have the same meaning in this paper and sample is the representation of population that research takes place.

Because the second method reflects features of population from statistical data, the regression model is

chosen for fetal estimation in this paper. Note, some terminologies such as *regression function*, *function*, *regression model*, *estimate function*, *estimate model* and *estimate formula* have the same meaning.

There are some estimate formula resulted from gestational researches such as [Hadlock 1985], [Duyet Phan 1985], [Nguyet Pham 2000], [Fusun Varol 2001], etc; some of them gain high accuracy but are only appropriate to population, community or ethnic group where such researches are done. If we apply these formulas into other community such as Vietnam, they are no longer accurate. Moreover, it is very difficult to find out a new and effective estimate formula or the cost of time and (computer) resources of formula discovery is expensive. Therefore, the first goal of this paper is to propose an effective algorithm which produces highly accurate formulas that are easy to tune with specified population. The process of producing formulas via such algorithm is as fast as possible. In addition, physicians and researchers always want to discover useful statistical information from measure sample and regression model. Thus, the second goal of this paper is to give facilities to physicians and researchers by introducing them a system or framework that implements such an effective algorithm in the first goal and builds up a tool allowing physicians and researchers to exploit and take advantage of useful and potential information under gestational sample. This tool is programmed as computer

software. In general, this paper has two objectives:

- Proposing an effective algorithm which produces highly accurate formulas. This algorithm is a heuristic approach that always results in optimal formulas by the fastest way.
- Introducing a framework that sets up the new algorithm in first goal and builds up a statistical tool which supports physicians and researchers in birth estimation domain. Moreover, physicians and researchers can discover new estimate formulas by themselves.

Section 2 gives an overview of the architecture of the framework. Section 3 is the description of effective algorithm producing the highly accurate formula. Section 4 discusses main use cases of framework with respect to gestational sample. Section 5 is the conclusion.

## 2. General Architecture of Framework

Based on clinical data input which includes fetal ultrasound measures such as *bpd*, *hc*, *ac*, *fl* and etc, the system produces optimal formulas for estimating fetal weight or fetal age with highest precision. Statistical information about fetal and gestation is also described in detailed in two forms: numerical format and graph format. So the framework consists of four components:

- *Dataset* component is responsible for managing information about fetal ultrasound measures such as *bpd*, *hc*, *ac*, *fl* and extra gestational information in reasonable and intelligent manner. This component allows other components to retrieve such information. Gestational information is organized into some abstract structure, e.g., a matrix whose each row represents a sample of *bpd*, *hc*, *ac*, *fl* measures. Following table is an example of this abstract structure:

Table 1. An example of gestational sample matrix.

<i>Bpd</i>	<i>Hc</i>	<i>fl</i>	<i>Ac</i>	<i>birth age</i>	<i>birth weight</i>
74	262	51	255	28	900
72	260	51	232	28	900
68	260	50	229	28	900
72	275	52	240	28	900
72	274	52	240	28	950
74	253	50	235	28	950
71	257	52	239	28	950
71	255	53	236	28	950
70	264	52	246	28	950

- *Regression model* component represents estimate formula or regression function. This component reads ultrasound information from *Dataset* component and builds up optimal estimate formula from such information. The algorithm used to discover and construct estimate formula is discussed in section 3. This component is the most important one because it implements such discovery algorithm.
- *Statistical manifest* component describes statistical information of both ultrasound measures and

regression function, for example: mean and standard deviation of *bpd* samples, sum of residuals and correlation coefficient of regression function, percentile graph of fetal weight. Statistical manifest is organized into two forms such as numerical format and graph format.

- *User interface (UI)* component is responsible for providing interaction between system and users such as physicians, researchers. A popular use case is that users enter ultrasound measures and requires system to print out both optimal estimate formula and statistical information about such ultrasound measures; moreover users can retrieve other information in *Dataset* component. *UI* component links to all of other components so as to give users as many facilities as possible.

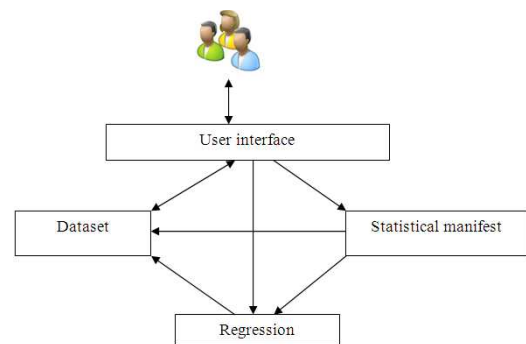


Figure 1. General architecture of framework.

Three components: *dataset*, *regression model* and *statistical manifest* are basic components. The fourth component is the bridge among them.

## 3. Algorithm Used in Framework

Suppose a regression function  $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n$  where  $Y$  is response or dependent variable and  $X_i$  (s) are regression or independent variables. Each  $\alpha_i$  is called regression coefficient. Response variable  $Y$  represents fetal weight or age. Regression variables  $X_i$  (s) are gestational ultrasound measures such as *bpd*, *hc*, *ac*, and *fl*. Given a set of measure values of  $X_i$  (s), the value of  $Y$  so-called *Y-estimate* calculated from this regression function is estimate fetal weight (or age) which is compared with real value of  $Y$  measured by ultrasonic machine. The real value of  $Y$  so-called *Y-real* is birth weight (or age). In this paper, the notation  $Y$  refers implicitly to *Y-estimate* if there is no explanation. The deviation between *Y-estimate* and *Y-real* is criterion used to assess the quality or the precision of regression function. The less this deviation is, the better regression function is. The goal of this paper is to find out the optimal regression function or estimate formula whose precision is highest.

A regression function will be good if it meets two conditions so-called:

- The correlation between *Y-estimate* and *Y-real* is large.
- The sum of residuals is small. Note that residual is

defined as the square of deviation between *Y-estimate* and *Y-real*,  $\text{residual} = (Y_{\text{estimate}} - Y_{\text{real}})^2$ .

These two conditions are called the *pair of optimal conditions*. A regression function is optimal or best if it satisfies the pair of optimal conditions at most, where correlation is largest and the sum of residuals is smallest. Given a set of regression variables  $X_i$  ( $i = \overline{1, n}$ ), we recognize that a regression function is a combination of  $k$  variables  $X_i$  (s) where  $k \leq n$  so that such combination achieves the pair of optimal conditions. Given a set of possible regression variables  $\text{VAR} = \{X_1, X_2, \dots, X_n\}$  being ultrasound measures, brute-force algorithm can be used to find out optimal function, which includes three following steps:

1. Let indicator number  $k$  is initialized 1, which responds to  $k$ -combination having  $k$  regression variables.
2. All combinations of  $n$  variables taken  $k$  are created. For each  $k$ -combination, the function built up by  $k$  variables in this  $k$ -combination is evaluated on the pair of optimal conditions; if such function satisfies these conditions then it is optimal function.
3. Indicator  $k$  is increased by 1. If  $k = n$  then algorithm stops, otherwise go back step 1.

The number of combinations which brute-force algorithm searches is:

$$\sum_{i=1}^n \frac{n!}{i!(n-i)!}$$

Where  $n$  is the number of regression variables. If  $n$  is large, there is the huge number of combinations, which causes the situation that algorithm never terminates and it is impossible to find out the best function. So we propose a new algorithm which overcomes this drawback and always find out the optimal function. In other words, the termination of new algorithm is determined and the time cost is decreased significantly because the searching space is reduced as small as possible. The new algorithm so-called heuristic algorithm is based on two assumptions about an optimal regression function which satisfies the pair of optimal conditions:

- First assumption: regression variables  $X_i$  (s) trends to be mutually independent. It means that any pair of  $X_i$  and  $X_j$  with  $i \neq j$  in an optimal function are mutually independent. The independence is reduced into the looser condition “the correlation coefficient of any pair of  $X_i$  and  $X_j$  is less than a threshold  $\delta$ ”. This is *minimum* assumption.
- Second assumption: each variable  $X_i$  contributes to the quality of optimal function. The concept of contribution rate of a variable  $X_i$  is defined as the correlation coefficient between such variable and *Y-real*. The higher contribution rate is, the more important respective variable is. Variables with higher contribution rate are called *high-contribute* variables. So optimal function includes only high-contribute regression variables. The second assumption is stated

that “the correlation coefficient of any regression variable  $X_i$  and real response value *Y-real* is greater than a threshold  $\varepsilon$ ”. This is *maximum* assumption.

The algorithm in this paper tries to find out a combination of regression variables  $X_i$  (s) so that such combination satisfies two above assumption. In other words, this combination constitutes an optimal regression function that satisfies two following conditions:

- The correlation coefficient of any pair of  $X_i$  and  $X_j$  is less than a minimum threshold  $\delta > 0$
- The correlation coefficient of any  $X_i$  and *Y-real* is greater than a maximum threshold  $\varepsilon > 0$

These two conditions are called the *pair of heuristic conditions*. Given a set of possible regression variables  $\text{VAR} = \{X_1, X_2, \dots, X_n\}$  being ultrasound measures, let  $f = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k$  ( $k \leq n$ ) be the estimate function and let  $\text{Re}(f) = \{X_1, X_2, \dots, X_n\}$  be its regression variables. Note that the value of  $f$  is fetal age or fetal weight.  $\text{Re}(f)$  is considered as the representation of  $f$ . Let *OPTIMAL* be the output of algorithm, which is a set of optimal functions returned. *OPTIMAL* is initialized as empty set. Let  $\text{Re}(\text{OPTIMAL})$  be a set of regression variables contained in all optimal functions  $f \in \text{OPTIMAL}$ . The algorithm includes four following steps:

1. Let  $C$  be the complement set of  $\text{VAR}$  with regard to *OPTIMAL*, we have  $C = \text{VAR} / \text{Re}(\text{OPTIMAL})$ .
2. Let  $G \subset C$  be a list of regression variables satisfying the pair of heuristic conditions. These variables are taken from complement set  $C$ . If  $G$  is empty, algorithm terminates; otherwise go to step 3.
3. We iterate over  $G$  in order to find out candidate list of good functions. For each regression variable  $X \in G$ , let  $L$  be the union set of optimal regression variables and  $X$ . We have  $L = \text{Re}(f) \cup \{X\}$  where  $f \in \text{OPTIMAL}$ . Suppose *CANDIDATE* is candidate list of good functions, which is initialized as empty set. Let  $g$  be the new function created from  $L$ ; in other words, regression variables of  $g$  belong to  $L$ ,  $\text{Re}(g) = L$ . If function  $g$  meets the pair of optimal conditions, it is added into *CANDIDATE*,  $\text{CANDIDATE} = \text{CANDIDATE} \cup \{g\}$ .
4. Let *BEST* be a set of best functions taken from *CANDIDATE*. In other words, these functions belong to *CANDIDATE* and satisfy the pair of optimal conditions at most, where correlation is largest and the sum of residuals is smallest. If *BEST* equals *OPTIMAL* then algorithm stops; otherwise assigning *BEST* to *OPTIMAL* and going back step 1. Note that two sets are equal if their elements are the same.

It is easy to recognize that the essence of algorithm is to reduce search space by choosing regression variables satisfying heuristic assumption as “seeds”. Optimal functions are composed of these seeds. Algorithm always delivers best functions but can lose other good functions. The length of function is defined as the number of its regression variables. The optimal bias is defined as the difference between two functions about correlation and sum

of residuals in optimal conditions. Terminate condition is that no more optimal functions can be found out or possible variables are browsed exhaustively. So the result function is the longest one but some other shorter functions may be optimal with insignificant optimal bias.

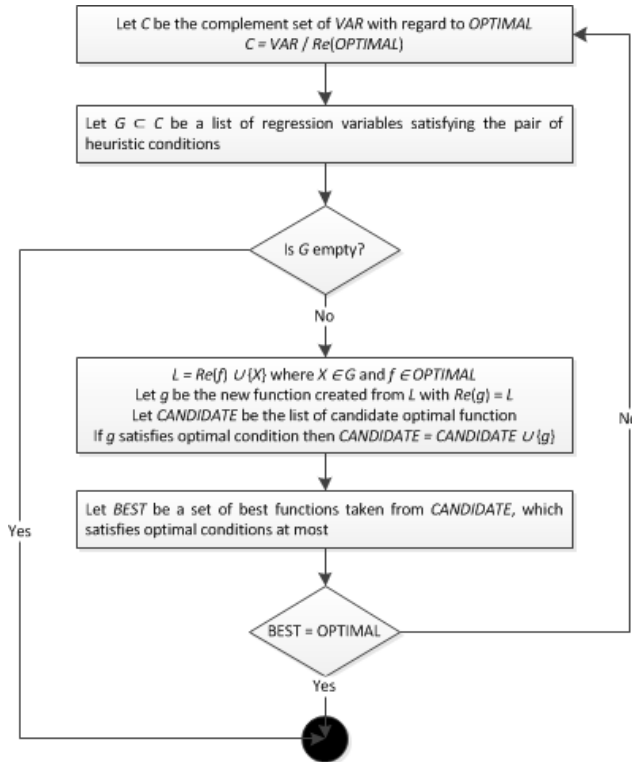


Figure 2. Heuristic algorithm flow chart.

## 4. Use Cases of Framework

The framework has three basic use cases realized by three components dataset, regression model and statistical manifest discussed in section 2. Three basic use cases includes:

- Discovering quality formulas with high accuracy. This use case is the result of algorithm in section 3.
- Providing statistical information under gestational sample. Statistical information is in numeric format and graph format.
- Comparison among different formulas.

### 4.1. Use case 1: Discovering Quality Formulas

Given gestational data [Hang Ho 2011] is composed of 2-dimension ultrasound measures of pregnant women. These women and their husbands are Vietnamese. These measures are taken at Vinh Long polyclinic, which include *bpd*, *hc*, *ac*, *fl*, *birth age* and *birth weight*. These women's periods are regular and their last period is determined. Each of them has only one alive fetus. Fetal age is from 28 weeks to 42 weeks. Delivery time is not over 48 hours since ultrasound scan. Gestational sample is shown in following figure.

No	bpd	hc	fl	ac	birth age	birth weight
1	74.0	262.0	51.0	255.0	28.0	900.0
2	72.0	260.0	51.0	232.0	28.0	900.0
3	68.0	260.0	50.0	229.0	28.0	900.0
4	72.0	275.0	52.0	240.0	28.0	900.0
5	72.0	274.0	52.0	240.0	28.0	950.0
6	74.0	253.0	50.0	235.0	28.0	950.0
7	71.0	257.0	52.0	239.0	28.0	950.0
8	71.0	255.0	53.0	236.0	28.0	950.0
9	70.0	264.0	52.0	246.0	28.0	950.0
10	70.0	248.0	51.0	229.0	28.0	960.0

Figure 3. Gestational sample.

After specifying minimum and maximum thresholds and which measures are regression variables and response variable, users will find out optimal formulas or functions as the results of algorithm in section 2. Optimal formulas that users discovery via using framework are shown in following figure.

No	Regression mod.	Type	R	Error mean (%)	Error sd (%)	Count
1	birth weight = 0...	product_log	0.9636...	0.4116896648	9.233508...	4
2	birth weight = 0...	product_log10	0.9636...	0.4116896651	9.233508...	4
3	birth weight = 52...	exp	0.9635...	0.4416218879	9.615743...	4
4	birth weight = 52...	10^	0.9635...	0.4416219097	9.615743...	4
5	birth weight = 0...	product_log	0.9631...	0.4182974697	9.304774...	3
6	birth weight = 0...	product_log10	0.9631...	0.4182974698	9.304774...	3

Figure 4. Optimal weight estimate formulas.

### 4.2. Use Case 2: Providing Statistical Information

Statistical information is classified into two groups: gestational information and estimate information:

- Gestational information contains statistical attributes about fetal measures, for example: mean, median and standard deviation of *bpd* distribution.
- Estimation information contains attributes about estimate model (formula), for example: correlation coefficient, sum of residuals and estimate error of estimate model (formula).



In representation, statistical information is described in two forms: numeric format and graph format.

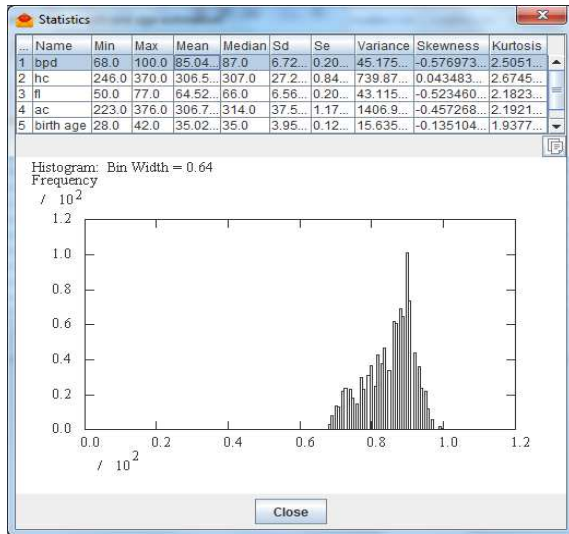


Figure 5. Gestational statistical information.

#### 4.3. Use Case 3: Comparison among Different Formulas

There are many criterions to evaluate the efficiency and accuracy of estimate formulas. These criterions are called evaluation criterions, for example: standard deviation, sum of residuals, estimate error, etc. Each formula has individual strong points and drawbacks. A formula is better than another one in terms of some criterions but may be worse than this other one in terms of different criterions. An optimal formula is the one that has more strong points than drawbacks in almost criterions. Hence, this framework supports the comparison among different formulas via criterion matrix represented in below figure. Each row in criterion matrix represents a formula whereas each column indicates the criterion. For example, first row, second row and third row represent formula in form of multiplication of logarithms, formula in form of exponent function and linear function, respectively. Three criterions: multivariate correlation, estimate correlation, estimate error and estimate ratio error are arranged in three respective columns.

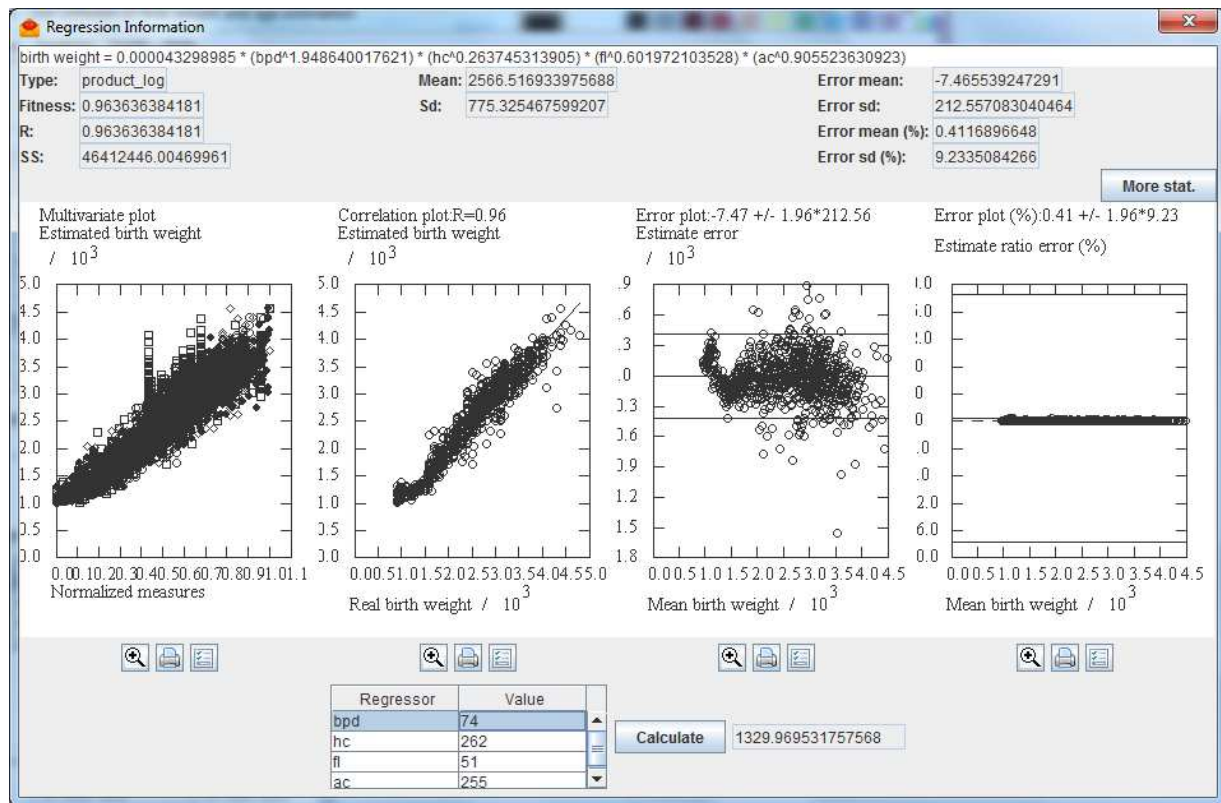


Figure 6. Estimate statistical information.

## 5. Conclusion

In general, this paper proposes the framework that gives scientists and physicians three utilities:

- Firstly, discovering new estimate formulas.
- Secondly, providing statistical information.
- Thirdly, comparison among different formulas based on pre-defined evaluation criterions.

Because the algorithm used to construct estimate formulas is based on heuristic assumptions, it gives optimal formulas but can lose other good formulas. In situation that scientists focus on some unusual criterion, such lost formulas are the ultimate for them but ignored. In the future, we improve this algorithm by adding constraints into heuristic assumptions. These constraints are made up of evaluation criterions and optimal formula considers both

the pair of heuristic conditions and these constraints. So, the architecture of framework is modified by adding a new component so-called evaluator component that manages

evaluation criterions and creates constraints from these criterions.

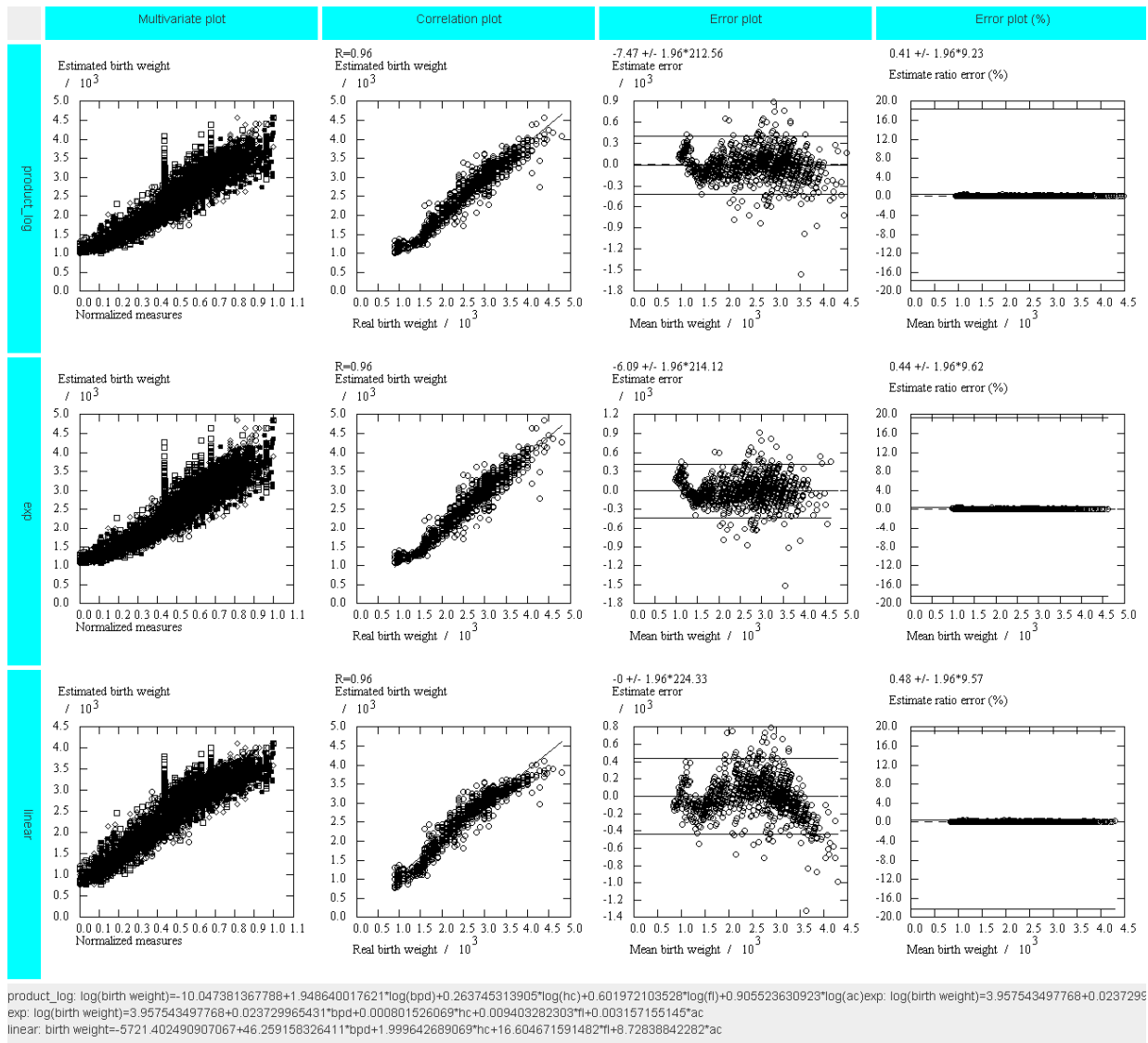


Figure 7. Comparison among different formulas.

## References

- [1] Hadlock FP, Harist RP, Sharman R (1985). Estimation of fetal weight with use of head, body and femur measurements: a prospective study. *Am J Obstet Gynaec* ; 21: 333-337.
- [2] Duyet Phan (1985). Ứng dụng siêu âm để chẩn đoán tuổi thai và cân nặng thai trong tử cung. Luận án Phó tiến sĩ y học, Trường Đại học Y Hà Nội.
- [3] Nguyen Pham (2000). Ước lượng cân nặng thai nhi qua các số đo của thai bằng siêu âm. Luận án tiến sĩ Y học, Trường Đại học Y Dược thành phố Hồ Chí Minh.
- [4] Fusun Varol, Ahmet Saltik, Petek Balkanlı Kaplan, Tulay Kilic and Turgut Yardim (2001). Evaluation of Gestational Age Based on Ultrasound Fetal Growth Measurements. *Yonsei Medical Journal*, vol. 42, No 3, pp. 299-303.
- [5] Hang Ho, Duyet Phan (2011). Ước lượng cân nặng của thai từ 37 – 42 tuần bằng siêu âm 2 chiều. Tạp chí Y học thực hành số 12 (797) năm 2011, tr. 8 - 9.