Physical Fitness Measurement using Multiple Linear Regression (Case Study of Children and Adolescents)

Zon Nyein Nway, Nan Saing Moon Kham
University Of Computer Studies, Yangon, Myanmar
zonnyeinnway7th@gmail.com, moonkhamucsy@gmail.com

Abstract

Physical activity is an essential component of a healthy lifestyle. It is clear that physically active people have a lower disease risk than sedentary individuals. Thus the measurement for physical fitness becomes the subject of current research interest. Fitness defines as the "The ability to carry out daily tasks with vigor and alertness, and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies." And Physical working capacity is the best measure of fitness. Therefore, a model is proposed to measure the physical fitness by estimating physical working capacity. By using this model, the value of fitness (PWC) can be predicted approximately. In this paper, the fitness value is calculated by the relational PWC which is the PWC value divided by the weight. The proposed fitness survey is implemented by the use of multiple linear regression to predict the physical fitness (PWC value) as the dependent variable by defining the relevant physical parameters such as lung capacity, resting heart rate, systolic blood pressure, diastolic blood pressure, breath holding, coordination of movement, weight and height ratio (WHR), motor response, shoulder girdle muscle (spin), flexibility of the spine, abdominal muscle (press) and squat as independent variables. The proposed model will give the physical activities which are the most related to the fitness and if these are made as regular practice, there will have more productivity.

Keywords: PWC (Physical Working Capacity), Weight height ratio (WHR).

1. Introduction

At present, more and more specialists in information technology are becoming active participants in solving the problems of human health. This means that information technology specialists have a sufficient amount of relevant medical knowledge to guide the development of information system related medical area. These attractive information systems contribute to the preservation of health in healthy people. Such systems are called "Navigators of Health". One of the key challenges in creating information systems such as "Heath Navigator" is the choice of parameters measured in the healthy person and processing the measurement results. Children and adolescents aged 8-to 17-year-old from No(2) basic education high school, Kyungyangone Township, Yangon, Myanmar have been performed as a case study this physical fitness measurement test[5] . In this multivariate analysis, physical working capacity was found to correlate significantly with age, lung capacity, resting heart rate, systolic blood pressure, diastolic blood pressure, flexibility of the spine (flexible), coordination of movement (cord), motor response (MR) , push-ups and press. No significant partial correlation was found between physical working capacity and breath holding and physical working capacity and weight height ratio (WHR). In the case study, setting PWC170/kg (as age groups are under 30) [7]) is set as dependent variable and other measuring variables are set as independent variables to regress the relationship between dependent and independent variables.

Physical health is critical for everybody. The population in Myanmar was reported at

59.98 million persons in 2009, according to the International Monetary Fund (IMF). In 2015, Myanmar's population is expected to be 67.59 million [9]. Due to this reason, it is very important to consider the fitness of citizens for the overall health of our country.

In 2008 and earlier, some developing countries such as Japan, Australia and China had implemented the fitness model by measuring physical working capacity and applied different variables, different methods and age-groups. Their models were best suited according to their environment and their countries which mean that those models were statistically significance for their countries. This proposed fitness model by measuring physical working capacity is distinctive in other models by considering new variables and different age-groups.

There are several studies concerning the tri-variate relation between physical work capacity, sex and age. In this study, several other variables were included simultaneously in the analysis: physical working capacity, lung capacity, resting heart rate, systolic blood pressure, diastolic blood pressure, breath holding, coordination of movement, weight and height ratio (WHR), motor response, shoulder girdle muscle (spin), flexibility of the spine, abdominal muscle (press) and squat. Approximately the total of 1200 children, males and females, aged 8 to 12 and the total of 1200 adolescents, males and females, aged 13 to 17 were included in this study.

3. Proposed System Framework

To implement the proposed fitness models (using the most related significant variables with physical working capacity), firstly the full fitness model (using all variables) must be built. The full fitness model is built by using multiple linear regression analysis method. And then the significant variables are selected by statistical t test. By using those significant variables, the stepwise multiple regression is redo in order to obtain the reduced fitness model. The Fisher's F test is used to test the significance of model. Then the Partial F test is used to compare the two models (full model and restricted model). To predict the physical fitness (as the dependent variable), the parameters such as lung capacity, resting heart rate, systolic blood pressure, diastolic blood pressure, breath holding, weight height ratio(WHR), flexibility of the spine (flexible), coordination of the movement (cord), motor response(MR), push-ups, press and squat test are used as independent variables.

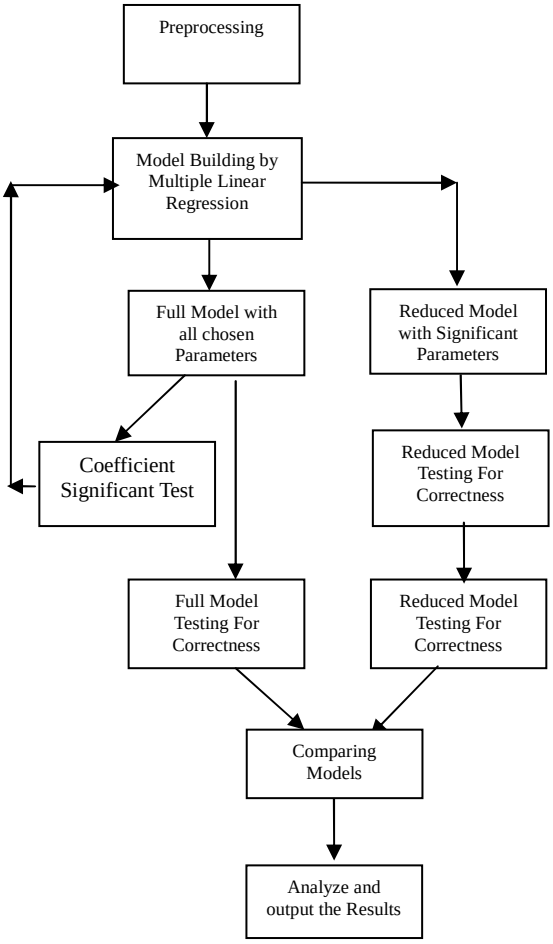

Figure1-Framework for Proposed Fitness Model

This work is developed the fitness models for all ages. And for adolescents the proposed model will give the physical activities which are most related to the fitness and make them regular practice, there will have more productivity in their working age than in adolescents. For adults, by making such related activities as regular practice, the diseases which can be caused by

lack of physical fitness such as heart disease, hypertension and diabetes can be reduced.

To evaluate the resulted model's significance, the model is tested based on Fisher's F-test and coefficients significant t-test. And the correctness of the model is evaluated by statistical assumptions such as autocorrelation, heteroscedasticity, residual analysis and normality test.

The measurable indicators of human morphological are the factors or independent variables (use 11 factors for children and 12 for adolescents in this system). Everyone has their own set of morphological parameters and its corresponding set of values for them.

3.1 Morphological and functional parameters

To predict the value of fitness, this system requires the measurements of physical activities. There are many fitness parameters that can measure physical performance. In this system, the 12 parameters shown in Table1 are chosen.

Table 1-Parameters for the regression analysis

Symbol	Symbol Definition
Response Y	PWC (Physical Performance)
X 1	Lung capacity (LC)
X 2	Resting Heart Rate(RHR)
X 3	Systolic Blood Pressure (BP-S)
X 4	Diastolic Blood Pressure(BP-D)
X 5	Breath Holding (BH)
X 6	Weight Height Ratio (WHR)
X 7	Coordination of Movement (Cord)
X 8	Flexibility of the Spine (Flexible)
X 9	Motor Response
X 10	Shoulder Girdle Muscles(Push-ups or Spin)
X 11	Abdominal Muscles(Press)
X 12	Squat Test (Squat)

This paper describes the building of the fitness models, and compares the original model with new models. Although multiple regression methods were applied to regress dependent and independent variables, there is also a consideration of statistical tests about model significance and coefficients significance. This proposed system based on multiple linear regression and statistical rules.

The objective of this work is to introduce a regression based prediction and forecasting system for physical fitness then compares the original model and new model. Least squared method is applied in the finding of coefficients. The second part is to analyze models' results by comparing the original model and new model.

3.2. The Choice of Parameters Measured

The following twelve parameters are learned from fitness and health literature and measured from school children by the groups of experts in the medical field.

1. Lung Capacity- This measurement is performed in a standing position with the help of dry portable, electronic or water Spirometer. Examinee is proposed to make the deepest breath, embrace mouthpiece lips and make a full soft exhale for 3-5 seconds. If the exhaled air is not passed by the device, his or her nose should be held with fingers. In doubtful cases, the values obtained (LC) should be re-test.

This indicator shows the potential for lung whose main oxygenation and removal of carbon dioxide, which is formed as a result of tissue respiration. With the lack of physical activity of human adult, lung capacity is reduced or poorly developed (during the growth of the organism). This can limit the oxygen supply with little physical load. Smoking impairs the diffusion capacity of the lungs, because nicotine affects a significant number of alveoli and cause bronchial obstruction.

2. Resting heart rate- This measurement is the number of cuts produced by the heart in 1 minute (beats / min). Heart rate measurement is performed by using the heart rate monitor showing the current value of the pulse rate. HR registration is made within 3 minutes after the

examinee has a sitting position, and fluctuations in heart rate are minimal. Minimum heart rate is recorded. The value of heart rate is measured in a state of relative peace of tension of the heart and cardiovascular system. This value depends on many factors, such as a person's age, the size of the heart, the condition of vessels and other factors.

3. BLOOD PRESSURE - This is measured in a sitting position, fixing the Tonometer cuff shoulder. The card recorded systolic and diastolic blood pressure in millimeters of mercury. Deviations from the age norm of blood pressure can be caused by various reasons and reflect, or tension of the cardiovascular system, or circulatory failure. Recurring at rest hypertension characterizes the lack of consistency of the heart and vascular system changes, as blood circulation, which can lead to persistent violations of blood pressure, and eventually to hypertension. The low blood pressure (among other reasons) can be caused by heart failure. The importance of blood pressure determines the inclusion of BP measurement in almost all the diagnostic programs on human health.

5. Weight Height Ratio (WHR) – This measurement is the ratio of body weight and height.

6. Breath Holding (BH) - Breath holding test is performed in a sitting position. Examinee is invited to perform consistently and arbitrarily maximum inhale and exhale, and then inhale for three quarters of the maximum and the maximum duration of hold your breath, hold your nose with your fingers. Duration of breath is recorded in the total number of seconds.

7. Shoulder girdle muscles (Push-ups or Spin) - This holds the maximum number of push-ups in 30 seconds. The subjects occupy a starting position in the emphasis lying on the chest, team-start push-ups, trying to do as much movement for 10 seconds, keeping the range of motion and preventing flexion (sagging) of the body. In the process of job counts the number of push-ups.

8. Press – This test is the maximum number of press for 30 seconds. This measurement is also called the test of abdominal muscles.

9. Squat Test- This is performed in the form of 30 sit-ups in 45 seconds. Tempo, duration and total number of sit-ups in the sample set ritmolider, which simplifies the standardization of the test. Squats are performed under strictly signals ritmolider job heels on the floor, from the starting position, standing feet shoulder width apart, with complete straightening the legs and back when lifting.

10. Motor Response (MR)-This is estimated to test the capture of the incident line (line length 40 cm). "Working" hand, is straight forward horizontally, vertically installed palm, thumb dismissed horizontally perpendicular to the ground and hands. The tester determines the line vertically, the zero mark on the upper level of the index finger. Examinee follows the fall line to capture the moment, to take it as soon as possible. The tester lets lineup in 1-5 seconds after the command "attention", varying duration of individual efforts. Record the distance from the top to the zero mark on the ruler.

11. Flexibility of the Spine (Flexible)- This is measured in the test with an inclination forward, from a position standing on the gym bench or step platform, latching fingers for two seconds to the lowest point possible on the measuring range. During this knee flexion is not allowed. If the examinee does not reach to zero, it is estimated the number of centimeters of flexibility with a minus sign.

12. Coordination of Movement (Cord)-This determined to test the throwing of the wall and bounced catching balls from a distance of 3 meters. Surveyed consistently throws 6 goals (3 times the right and left hand 3 times) into the wall at 2-3 meters from the floor and catches the balls bounce. Motor coordination describes the ability to perform complex coordination of movements.

3.3 Background Theory: Multiple Linear Regression Analysis

In regression analysis, there is a dependent variable which is trying to explain, and one or more independent variables that are related to it. This framework allows you to safely bring to the process of obtaining estimates of regression coefficients and to carry out basic statistical testing of hypotheses. The response (PWC) is associated with morphological and functional parameters (factors) by the following relation:

$$Y=B_0+B_1X_1+B_2X_2+...+B_nX_n+\in \quad (1)$$

Where,
Y=a response (dependent variable)
B_0= evaluation of the constant component
$B_i=i^{th}$ component of multiple regression
(i=0,1,..,n)
$X_i=i^{th}$ independent variable,
$\in$ =error.

The task of building the approximation of multivariate regression models can be considered complete when the results of experiments obtained by estimating the coefficients of equation (1) such as B_0, B_1,…,B_n.

We can calculate the coefficients by the formula (2).

$$B=(X'X)^{-1}X'Y \quad \text{…….. (2)}$$

Where,
X'= The transposed matrix of independent variables
$(X`X)^{-1}$ =Inverse of the product of the transposed matrix of independent variables on the very matrix of independent variables
Y = Vector of observations

The experiment will involve almost 2000 and above: children (males and females aged 8 to 12 years) and adolescents (males and females aged 13 to 17 years).Therefore, the regression equations are calculate for all samples, and sampled by sex and age.

3.4 System Evaluation

The multiple regression equation can be calculated from the coefficients b0…b12 by the formula (2). The analysis was made in order to test whether each of the obtained coefficients of the regression model statistically significant or not. To answer this question, we use the following parameters: the standard error of each of the regression coefficients, t-statistics (t-stat) and P-value.

To test the hypotheses of the significance of the coefficient we use the student t-test. In this work, using t-test compared to the value of t-statistics from the value of t in a T-table. If the calculated t-value is greater than or equal to the critical t-value from the T-table, the regression coefficient is statistically significant at the chosen significance level. In hypothesis testing, significance level 95% is recommended.

Now, using analysis of variance, we must analyze whether the resulting model is useful or not. The analysis of variance reveals the dependence of the experimental data by examining the significance of differences in mean values.

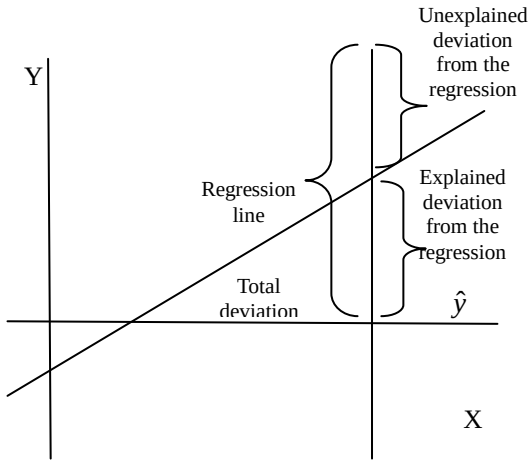

Figure 2- The sample variance for the regression model

For the case of multiple linear regression, use the concept of total variance, explained variance and unexplained variance (see Figure 1). To clarify the usefulness of the regression line, the values of the F-statistic is used. F-statistic is the ratio of explained variance (mean square regression), and unexplained variance

(standard error). If the resulting value of the F-statistic is greater than unity, the regression line is useful.

To test the significance of the model, we use the F distribution table. The use of F-criterion is compared to the value of the F-statistic with the value of F-test in the F-table. If the calculated value of the F-statistic is greater than or equal to the critical value of F-test in the F-table, we can conclude that the regression model is statistically significant.

Therefore, if the value of F statistic greater than the value of F-criteria in the F-table, the equation of the regression model is statistically significant, meaning that the regression model is useful.

To quantify how useful the resulting multivariable model of predicting the PWC, we apply the coefficient of determination R^2.

But the use of R^2 to access the quality of model has the disadvantage that the inclusion in the model of the new factor automatically increases the value of R2. Therefore, when a large number of factors are used, it is preferable to use the adjusted coefficient of multiple determinations $R^2_{adj.}$ The determination coefficient R^2 and the adjusted coefficient of determination R^2_{adj} can be calculated as followed.

$$R^2_{adj} = 1 - ((\frac{n-1}{n-k-1})(1-R^2)) \quad \text{.....(3)}$$

Here,

y_i = i-th value of the dependent variable, PWC,

$\hat{y}_i$ = i-th predicted value of PWC,

k = the mean value of the dependent variable, PWC,

k = the number of regressors,

n = number of observations

We estimate the quality of the model using the standard error of estimate SEE, calculated by the formula.

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}} \text{.} \quad \text{..... (4)}$$

3.5 Objectives of the Proposed System

The main purpose of this dissertation is to design fitness models, which can predict the physical fitness of people. The research motto is that “intends to design a fitness model that based on Fisher's F-test” .The major objectives of the proposed fitness models are:

(1)To determine which indexes are best measures for relative physical working capacity

(2)To analyze physical working capacity of Myanmar people

(3)To develop mathematical fitness models

(4)To compute the relevant parameters for physical fitness by applying the methods of multiple-regression analysis and statistical rules

By implementing this fitness models, This model can be applied in gyms, schools and government training centers and even in factories of workshop to examine the physical fitness. And this PWC models can also be applied in all government organizations (such as garments, fishery, steel factory, civil service training center, etc) and private economic organizations where numerical strength of work force is required. Some work places where there may be many risks and there may be required many work force (such as construction groups, private companies and the training centers with camp, military, NGO & UNICEPT in Government organizations), physical working capacity is also important as well as degrees of qualifications.

4. Experimental Results for Case Study

This work describes model building for physical fitness, comparing the full model’s accuracy (based all of the considered parameters) and the reduced model’s accuracy (based only significant parameters). By the case study, the collected data is different types of physical

Table2. Comparison for Model Accuracy and Correctness for the Case Study

Models by age	R^2_a	The Standard Error of Estimate	Average value of predicted PWC
8 Females	0.216	1.9331	9.6408
9 Females	0.059	2.136	9.8467
10 Females	0.192	2.0365	9.6875
11 Females	0.141	2.11	10.11083
12 Females	0.194	2.09	10.0217
13 Females	0.026	0.23651	10.0933
8 Males	0.134	2.289	10.6825
9 Males	0.198	2.18418	10.41833
10 Males	0.161	2.161	10.7383
11 Males	0.066	2.245	10.9217
12 Males	0.113	2.553	10.7983
13 Males	0.224	2.406	10.9933
Males 8to13	0.1062	2.3718	10.7447
Females8to13	0.1035	2.1783	9.9001
M+F(8to13)	0.0985	2.3199	10.32944
15 Females	0.291	2.06002	9.5883
16 Females	0.271	0.21448	9.8892
17 Females	0.239	2.0926	9.7833
14 Males	0.236	2.51178	11.175
15 Males	0.147	2.442	11.4833
16 Males	0.328	2.416	11.7867
17 Males	0.333	2.229	11.6667
Males14to17	0.2736	2.389	11.5279
Females14to17	0.2312	2.1480	9.7040
M+F(14to17)	0.2804	2.3621	10.6159

activities those are also suitable for physical fitness of adolescents at the age of eight to seventeen year old. The next stage is finding coefficient values for each parameter by least square method
and building regression models by multiple linear regression and testing the models by applying statistical rules.

Table.2 describes the models' correctness based on multiple R, coefficient of determination R^2, adjusted R^2 and standard error of estimate. The standard error of the estimate is a measure of the accuracy of predictions. The regression line is the line that minimizes the sum of squared deviations of prediction (also call the sum of squares error), and the standard error of the estimate is the square root of the average squared deviation. Therefore, in a regression line, the smaller the standard error of the estimate is, the more accurate the predictions are. By comparing the value of the standard error estimates with the average value of the predicted values of the dependent variable, the standard error of the estimate is not greater than the average value.

The values of the coefficients of determination (R squared) are low for all models from eight to thirteen year old males and females. But the F-statistics' value for all models are statistically significant. In this case, the value of R is intended for the size of a square sample, and a value of the F-statistic is for the intended size of the population. So, the model may be useful for predicting and not have to worry about R-squared.

In each model from eight to thirteen year old males and females, eleven types of physical activities are used in the regression system. But for each model from fourteen to seventeen year old males and females, twelve types of physical activities are used in the regression system.

In the analysis for eight to seventeen year old females and males,

1. All independent variables have a linear relationship with the dependent variable for all models, but there is a weak relationship between dependent and independent parameter settings.
2. Multicollinearity is absent in all models.
3. Compared model (partial test F), all of the reduced models are better than full model.

4. All models do not suffer from heteroscedasticity.
5. All models do not suffer from autocorrelation by the test of Durbin Watson.
6. The values of the coefficients of determination (R squared) are low for all models from eight to thirteen year old females and males. But the F-statistic for all models is statistically significant. In this case, the value of R is intended for the size of a square sample, and a value of the F-statistic for the intended size of the population. So, the model may be useful for prediction and not have to worry about R- squared.
7. Heart rate indicator exists in all models except the eleven and twelve males. Therefore, we can assume that this option is the best indicator for predicting physical performance between the ages of eight to thirteen year old males and females.
8. Indicator test squat exists in all models for fourteen to seventeen year old males and females. Therefore, this option is the best indicator for predicting physical performance at the age of fourteen to seventeen year old males and females.
9. All regression models are statistically significant, it certainly does not explain a lot of variations by the results of adjusted coefficient of determination ($adjR^2$), and it is likely that the models will predict future observations with high accuracy because the standard error of estimate is not greater than the average of predicted PWC value.

4. System Discussion

The physical health and fitness model (by setting physical working capacity as a dependent variable) based on eleven and twelve measurements of physical activities by using multiple linear regression has been proposed. These models showed the significant parameters (physical activities) to be applied in the population for performing fitness assessment for people. By analyzing the evaluated significant parameters, the suitable physical activities for each age group can be determined in order to become fit and healthy people.

To predict the physical health and fitness of people, selected method of multiple regression analysis of statistics is used. The mathematical statistical models are conducted to predict the physical health of people of any age. But the case study is between the ages of eight to seventeen years. The parameters (different types of physical activities) for the mathematical fitness prediction models are considered. With which the physical health and fitness for people can be quickly and efficiently assessed. Regression analysis is used to regress the considered physical activities and dependent parameter PWC. And the statistical methods and rules are applied to predict whether the resulted models were statistically significant or not in order to use in assessing the health status. And as a case study, in order to predict the physical health of adolescents between the ages of eight to seventeen, significant variables associated with fitness are evaluated and the physical health of adolescents.

6. Conclusion

The resulted PWC model is well suited to test the physical fitness. This system predicts the physical health of children and adolescents as a case study by selected method of multiple regression analysis of statistics, which allows the analysis of multivariate statistical models. The mathematical statistical models are conducted to predict the physical fitness of adolescents between the ages of eight to seventeen years. The eleven parameters (different types of physical activities) for 8 to 13-year-old males and females and twelve parameters for 14 to 17-year-old males and females are considered. Regression analysis is used to regress the considered physical activities and dependent parameter PWC. And the statistical methods and rules are applied to explore whether the resulted models were statistically significant or not in order to use in assessing the health status. Significant parameters are found for evaluation the physical fitness of children and adolescents between the ages of eight to seventeen years as a case study.

References

[1]http://en.wikipedia.org/wiki/Central_limit_theorem
[2]http://crow.academy.ru/econometrics/lectures_/lect_04_/demo_04_/sld031.htm
[3]http://en.wikipedia.org/wiki/Regression_analysis
[4]http://en.wikipedia.org/wiki/Linear_regression
[5]http://faculty.chass.ncsu.edu/garson/PA765/regress.htm
[6] B.G. Jonson and I. Astrand, Physical Work Capacity in Men and Women Aged 18 to 65, Scand J Soc Med 7: 131-142, 1979.
[7] H.A. Devries, G.R. Brodowicz, L.D. Robertson, M.D. Svoboda, J.S. Schendel. A.M. Tichy and M.W. Tichy, Estimating Physical Working Capacity and Training Changes in the elderly at the fatigue threshold (PWC ft), Ergonomics, 1989, Vol.32, No.8, 967-977.
[8] H.A. Devries, M.W. Tichy, T.J.Housh, K.D.Smyth, A.M. Tichy and D.J.Housh, A method for estimating physical working capacity at the fatigue threshold (PWC ft), Erogonomics, 1987, vol.30, No.8, 1195-1204.
[9] H.A. Devries, T.J. Housh, G.O. Johnson, S.A. Evans, G.D. Tharp, D.J. Housh and R.A. Hughes, Factors affecting the estimation of physical working capacity at the fatigue threshold, Ergonomics, 1990, vol.33, No.1, 25-3.
[10] http://pareonline.net/getvn.asp?v=8&n=2
[11]http://explorable.com/calculate-standard-deviation.html
[12]http://sonia.hubpages.com/hub/stddev
[13]http://www.mathsrevision.net/gcse/pages.php?page=42
[14]http://en.wikipedia.org/wiki/Linear_regression
[15]http://faculty.chass.ncsu.edu/garson/PA765/regress.htm
[16] B.G. Jonson and I. Astrand, Physical Work Capacity in Men and Women Aged 18 to 65, Scand J Soc Med 7: 131-142, 1979.
[17] Professor Myint Han: Health Care of Elderly in Myanmar, Regional Health Forum-Vol 16, Number 1, 2012.
[18]Bj€orn Wohlfart and Gholam R. Farazdaghi : Reference values for the physical work capacity on a bicycle ergometer for men – a comparison with a previous study on women , Department of Clinical Physiology, Lund University Hospital, Sweden