Analysis of Physical Health by Estimating Physical Working Capacity

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

Abstract

It is clear that physically active people have a lower disease risk than the others who don't have. Nowadays, the components of activity which determine particular health gains are poorly understood. Thus the analysis for physical health is the subject of current research interest. In this paper, a model is proposed to analyze the physical health by estimating physical working capacity. This study shows associations between physical parameters and physical performance by using multiple linear regressions. The aim of the use of multiple linear regression is to minimize the error in our prediction of the dependent variable, and minimize the residuals we made. And regression is a statistical method and is the most suitable method for this statistical information system. This paper is concerned with future contributions to a physical performance for exercise recommendations both to the public and to individuals.

1. Introduction

At present, 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 person and the development of algorithms for processing measurement results. Among all of the variety of possible solutions to this problem, this paper would like to select the best option and use it. The criteria for selection of parameters measured, the measurement and calculation of individual parameters can be the ease of measurement and measurement reliability, sensitivity and reproducibility. These measured parameters and test procedures used in this survey as children at the age of 13 to 17 and 22 to 35 years old people. After the measurement of these variables, need to determine which ones are most important for prediction. In this case, we determine the physical working capacity by analyzing the predictors. By examining the PWC, we can predict physical health. So as a dependent parameter, this system takes the physical fitness of man (PWC). As setting PWC (Physical Working Capacity, PWC also known as PWC 170/kg) is one of the most important components of physical health..

2. Related Work

Analysis on physical health becomes a popular topic recently. Most research trends focus on regression analysis to analyze physical health. A preliminary study of ……….was reported in [8]. A method for estimating physical working capacity at the fatigue threshold presented that presently available tests of physical working capacity (PWC) such as VO2 max and critical power may not be appropriate for unfit subjects because they require maximal or supra-maximal workloads. Therefore, the purpose of this study was to evaluate a sub-maximal discontinuous incremental bicycle ergo meter test (PWCFT) with an end point (fatigue threshold) determined by recording electro-myographic (EMG) fatigue curves in the quadriceps muscle.

Physical Work Capacity in men and women aged 18 to 65 is presented in [6]. A probability sample of the population of a region in central Sweden performed a sub-maximal exercise test on bicycle ergo meters. The participants were men and women aged 18 to 65 taking part in the REBUS study. In multivariate analysis, heart rate at the rate of work 100Wand 50W was found to correlate significantly with sex , body weight and stature, blood pressure at rest, physical training in leisure time, and physical strain at work. No significant partial correlation was found between heart rate a work and alcohol consumption.

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, vital capacity, pulse at rest, systolic blood pressure, diastolic blood pressure, breath, BMI, flexibility of the spine, visual motor response, shoulder girdle muscles, abdominal muscles, sit to stand, sit to reach, body weight, height, chest, hip, waist, waist/hip, squat test, vertical jump. Approximately 250 of the total of 1000 individuals, boys and girls, aged 13 to 17 and 26 to 35 in the selected area were included in this study.

3. Proposed System Framework

Physical fitness is not only one of the most important keys to a healthy body but also it is the basic of dynamic and creative intellectual activity. The level of overall muscle endurance capacity is determined such as respiratory, cardiovascular, nervous and endocrine systems, coordination of their work during exercise, and ultimately can serve as a generalized assessment of the physical condition of the body. In this system, after the measurement of morphological and functional indicators of physical health, we plan to build forecast values of

PWC using multiple regressions. And then this system will analyze the results. In this case, it is necessary to find a mathematical relationship of human physical performance measured by morphological and functional parameters.

The measurable indicators of human morphological are the factors or independent variables (we use 18 factors in our case. Everyone has their own set of morphological parameters and its corresponding set of values for them. The factors are deterministic, as obtained by independent measurements.

3.1 Morphological and functional parameters

To predict the value of physical working capacity, this system requires the measurements of physical activities. There are many fitness parameters that can measure physical performance. In this system, the following 18 parameters are chosen. Among them, hypothesis test, F test, T test and residual normality are used to select the successful parameters.

Table 1-Parameters for the regression analysis

Symbol	Symbol Definition
Response Y	PWC (Physical Performance)
X 1	Vital capacity
X 2	Pulse at rest(heart rate, beat/min)
X 3	Systolic Blood Pressure
X 4	Diastolic Blood Pressure
X 5	Breath
X 6	BMI (Body Mass Index)
X 7	Flexibility of the Spine
X 8	Visual motor response
X 9	Shoulder Girdle Muscles(Push-ups)
X 10	Abdominal Muscles(Sit-ups)
X 11	Sit to stand
X 12	Sit to reach
X 13	Body Weight
X 14	Height
X 15	Chest
X 16	Hip
X 17	Waist
X 18	Waist/Hip

In this case, the analysis uses the following 18 morphological and functional parameters for a group of girls and boys aged between 13 and 17 and 25 to 35 years.

3.2. The Choice of Parameters Measured

In this system, we use 18 parameters as shown in Table 1. Some are easy to measure but some aren't. Therefore, some difficult parameters are explained.

1. Weight- measured in light clothes and without shoes. The measurement results are recorded in the "map" in kg.

2. Resting heart rate- Heart rate (HR) and pulse rate- the number of cuts produced by the heart in 1 minute (beat/min). Measurement of heart rate produced by heart rate monitor shows an integrated, current heart rate. HR registration is made within 3 minutes after the examinee has a sitting position, and fluctuations in heart rate are minimal.

3. BLOOD PRESSURE - measured in a sitting position. In the "map" recorded systolic and diastolic blood pressure in millimeters of mercury. Blood pressure (BP) is a key measure of the function of the cardiovascular system.

5. Vital capacity (VC) - measurement is performed in a standing position with the help of electronic, portable dry or water spirometer. Examinee is proposed to make the most deep breath, embrace mouthpiece lips and make a full soft breath for 3-5 seconds. To part of the exhaled air is not passed by the device should hold his nose with your fingers or a clamp. In doubtful cases, the values obtained (VC) should be made to re-test. In the "map" of the readings recorded in milliliters. VC is an important indicator of the functionality of the respiratory system and depends on the total volume of the lungs, respiratory muscle strength, resistance of the chest and lungs, their tension, wears off and other factors.

6. Visual-motor response - is estimated to test the capture of the incident line (line length 40 cm). "Working" hand, is straight forward grizontalno, palm mounted vertically, thumb dismissed perpendicular to the palm. The tester determines the line vertically, the zero mark on the upper level of the index finger. After fixing the position of the hands and the line installation, the command attention must be given. 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 there's room to the zero mark on the ruler. The results expressed in centimeters, is entered in the "map". This indicator measures the number of centimeters for the fall, describes the mobility of nervous processes.

7. Shoulder girdle muscles - holds the maximum number of push-ups in 10 seconds. Boys - relying toes (girls - knee), and palms of hands. In the process of job counts the number of push-ups, and the result is stored in the appropriate position, "Maps".

8. Sit to reach and sit to stand - The Sit and Reach Test is the most commonly used method of assessing flexibility. This test measures the flexibility of the lower back and hamstring muscles. Required equipment include either a sit and reach box, a ruler, or some other instrument for measuring.

9. Abdominal muscles (Sit ups) -The sit-up is a callisthenic abdominal exercise that works the rectus abdominus muscles. It was once considered the gold standard for working the midsection, but fell out of favor due to controversy dealing with lower back injuries. The sit-up is still used by many military personnel as a gauge of abdominal endurance.

3.3 Background Theory of Regression Analysis

In regression analysis, there is a dependent variable which is the one you are 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 about the equation of the regression coefficients and to carry out basic statistical testing of hypothesis, we can assume that 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=ith component of multiple regression (i=0,1,..,n)
X_i=ith independent variable,
E=error.

Now the equation (1) can be written for any response in the form:

$$\hat{y}_i = b_0 + b_1x_1 + b_2x_2 + ... + b_nx_n + \in_i \quad (2)$$

From which it follows that for any set of X values corresponding to the Y-value is the value of $b_0+b_1x_1+b_2x_2+…+b_nx_n+$ e, taking into account that any individual Y is able to avoid regression to the surface, which embodies a set of mean values of Y.

The task of building a first approximation of multivariate regression models can be considered complete when the results of experiments obtained by estimating the coefficients of equation (2) such as b_0, $b_1,…,b_n$.

We can calculate the coefficient $b_0…b_6$ by the formula (3).

$$B=(X'X)^{-1}X'Y \quad (3)$$

According to the procedure of multiple regression analysis, it is first necessary to know each of the independent variables. It is correlated with the dependent parameter PWC. To assess this, correlation is proposed by using linear regression.

The experiment will involve almost 2500 and above: boys and girls aged 13 to 17 years and 25 to 35 years. Therefore, the regression equation must have to calculate for all samples, separately for boys and girls.

Now we are presenting the results of analysis for the 250 samples by using the seven morphological parameters.

3.4 Results of the analysis

After examining the seven morphological parameters, we must consider whether the parameters are linearly dependent or not. To answer this question, there may be required the scatter plots for each relationship. In every scatter plots, we will see the slope equations, R2 (the value of the coefficient of determination). If the value of R2 is equal to zero, we can conclude that parameter is linearly independent. The best result may be the value that greater than zero and not greater than one.

Table 2-Parameters for the regression analysis

Symbols	Symbol Definition
X1	Body weight
X2	Height
X3	Chest
X4	Waist
X5	Hip
X6	Shoulder
X7	Waist/Hip

In Table 3, we can see that the effect of each of the 7 listed morphological and functional parameters can be described as a linear or weak. If each parameter is linearly independent, it means that that parameter is usefulness and we must remove it.

Table 3. The linear dependence of the parameters number

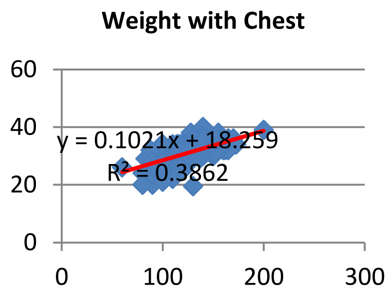
No	Parameters	Linearly dependence or not
1	Body weight	Yes
2	Height	Yes
3	Chest	Yes
4	Waist	Yes
5	Hip	Yes
6	Shoulder	Yes
7	Waist/Hip	Yes

Then we can calculate a correlation matrix between the above morphological parameters. The result is shown in the following table.

Table 4. The Correlation Matrix on Measured Parameters

	Weight	*Height(Inches)*	*Chest*	*Waist*	*Hip*	*Shoulder*	*Waist/Hip*
Weight	1						
Height(Inches)	0.547922	1					
Chest	0.621411	0.197259	1				
Waist	0.74786	0.399486	0.77242	1			
Hip	0.685259	0.3039	0.862353	0.752272	1		
Shoulder	0.535995	0.600592	0.165528	0.414613	0.246018	1	
Waist/Hip	0.250772	0.177909	0.105598	0.558859	-0.10209	0.275623	1

The following are the scatter plots for each parameter and dependent variable (PWC).



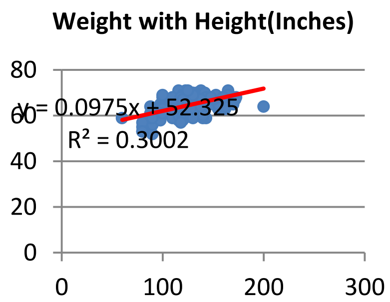


Figure 1. Relation between dependent and independent variables

We can calculate the coefficients $b_0 \ldots b_6$ by the formula (3) and the results are listed in Table (5).

Table 5. The coefficients of regression analysis

b_0	-74.5089498
b_1	1.094112462
b_2	0.262977208
b_3	2.642004399
b_4	0.927077433
b_5	1.927408477
b_6	-25.05642779

The analysis was made in order to obtain an answer to the question whether each of the obtained coefficients of the regression model is 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.

These parameters are important for testing the significant of each of the coefficients. To test the hypothesis that actually need to know the value t-statistic for each factor, as t-statistics

(t-stat) allows you to test the significance of each of the regression coefficients.

In particular, the t- statistics for each of the coefficients can be calculated using the standard errors of regression coefficients.

To calculate the standard errors of the coefficients b0…b6 uses the following formula.

$$SE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \Big/ (n-k-1)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \quad (5)$$

$\hat{y}$ Where

SE- standard error of regression coefficient,

yi - measured values PWC, (dependent variable)

$\hat{y}$ i[th] predicted value of PWC

xi -i[th] independent variable

i= 1…..n,

x bar - average value of the independent variables

k - Regression degrees of freedom or the number of repressors,

n - Number of observations

You can then calculate the t-statistic (t-stat) for each factor using the following formula

$$t(b) = \frac{b}{SE_b} \quad (6)$$

Where

T (b)-t-statistics of one factor

b -regression coefficient

SE-standard error of regression coefficient

Table 6. The calculated t-statistic (t-stat) of each regression coefficients

Regression coefficients	Standard error	T-statistic (t-stat)
b 0	38.73642661	-1.92349
b 1	0.269052877	4.066533
b 2	0.498437089	0.527604
b 3	1.348408177	1.959351
b 4	1.053970885	0.879604
b 5	0.471474684	4.088042
b 6	45.14598923	-0.579428

To test the hypotheses, we use the significance of the coefficients. In our case, using t-test (T-test) compared to the value of t-statistics from the value of T-test in a T-table.

If the calculated t-value greater than or equal to the critical value T in the T-table, we can conclude that the regression coefficient is statistically significant.

If the t-statistics in absolute value greater than T-test, the coefficient is significant at the chosen significance level. In hypothesis testing, significance level is the criterion of rejection of the null hypothesis. Thus, if you choose the significance level =10%.

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

For the case of linear regression, using the concept of total variance explained variance and unexplained variance (see Figure 2), we can say that the resulting regression line in the experiment is useful if it significantly decreases the unexplained variance.

To clarify the usefulness of the regression line, we use the values of the F-statistic. 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.

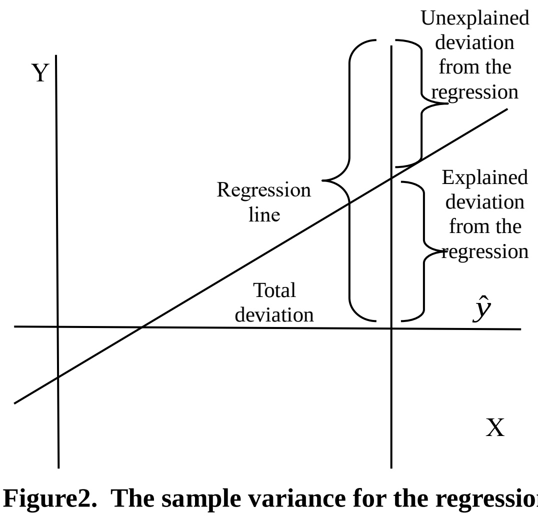


Figure2. The sample variance for the regression model

In this case, the explained variance or mean-square regression (MSR-Mean Squared Regression) can be calculated using the following formula.

$$MSR = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{k} \qquad (7)$$

Unexplained variance or mean square error (MSE-Mean Squared Error) is calculated by the formula.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}, \qquad (8)$$

Here,

MSR - RMS regression (explained variance), MSE-mean-square error (unexplained variance), y_i - i^th^ dependent variable PWC,

$\hat{y}_i$ - i^th^ predicted value - mean value of the dependent variable,

k - the number of regressors , n - number of observation

The value of F-statistics can be calculated using the following formula [9-10]:

$$F = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 \Big/ k}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \Big/ (n-k-1)} \qquad (9)$$

(Or)

$$F = \frac{MSR}{MSE}. \qquad (10)$$

To test the significance of the model, we use the F distribution table. In our case, the use of F-criterion is reduced 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, we can say that the equation of the regression model is statistically significant, meaning that the regression model is useful.

Let's try to quantify how useful the resulting multivariable model to predict the parameter PWC. In other words, what part of the value of the parameter prediction substantiates multifactor model of the PWC [11]. For quantitative estimation of this utility of the multivariate model, an attempt was made to use the coefficient of determination R2 [11]. In order to be able to compare models with different number of factors so that the number of factors (covariates) had no effect on the value commonly used adjusted coefficient of determination. The determination coefficient R2 and the adjusted coefficient of determination R^2_{adj} can be calculated as follows:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{11}$$

$$R^2{}_{adj} = 1 - ((\frac{n-1}{n-k-1})(1 - R^2)) \tag{12}$$

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.

$$SEE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}}. \tag{13}$$

4. Discussion

Path analysis is an extension of multiple regressions. Its aim is to provide estimates of the magnitude and significance of hypothesized causal connections between sets of variables. Therefore, path analysis can be used instead of multiple regressions and compared the analysis results. Furthermore, there are many procedures for estimating the shrinkage of R2 such as adjusted R2, cross-validation, data splitting and formula based estimation of the cross-validity coefficient. Depending on the pros and cons of each method, the analysis result may be altered. In every predictive research, the main emphasis is on practical applications.

5. Conclusion

The resulted PWC model is well suited to test the physical performance. This system firstly examines possible physical parameters related to physical working capacity. Secondly, brings to the obtaining estimates of regression coefficients and then an analysis was made in order to obtain an answer to the question whether each of the obtained coefficients of the regression model is statistically significant or not. Finally, using the analysis of variance, can analyze whether the resulting model is useful or not.

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