УДК 004.4, 519.2
05.13.17 – Теоретические основы информатики

Куи Тар Со
ФГБОУ ВПО “ Московский Государственный Технический Университет им. Н.Э. Баумана”
Россия, Москва[1]
Аспирант кафедры “Информационные системы и телекоммуникации”
E-Mail: kyithar82@gmail.com

Разработка математических моделей и программного обеспечения для физического здоровья человека

Аннотация. В статье представлены результаты исследований по созданию математической модели физического здоровья «здорового человека» методом регрессионного анализа, где в качестве факторов выступают физические параметры человека, а в качестве отклика – показатель физической работоспособности. Уравнение регрессии строилось для каждой возрастной группы. После выполнения множественного регрессионного анализа получена множественная регрессионная модель, которая может предсказать физическую работоспособность для девушек в возрасте от четырнадцати до семнадцати лет. В статье приведены результаты анализа и разработка регрессионной модели для пятнадцатилетних девушек. Также представлены проверки гипотез для модели, т. е проверка значимости модели, значимости коэффициентов, гетероскедастичности, автокорреляции, мультиколлинеарности и нормальности. В результате для прогнозирования физического здоровья человека выбран метод множественного регрессионного анализа статистики, который позволяет проводить анализ многофакторных статистических моделей. Разработаны математические модели и программное обеспечение для прогнозирования физического здоровья девушек в возрасте от четырнадцати до семнадцати лет. Определены значимые параметры для математических моделей прогнозирования, с помощью которых быстро и эффективно можно оценить физическое здоровье девушек в возрасте от четырнадцати до семнадцати лет.

Ключевые слова: регрессионная модель; корреляция; дисперсионный анализ; т – статистика; F- статистика; коэффициент детерминации; гетероскедастичность.

Идентификационный номер статьи в журнале 37TVN314

[1] 105005 г. Москва, 2-я Бауманская ул., д. 5, стр. 1

Введение

Необходимость количественного измерения соматического здоровья, оценки его резервов приобретает особую актуальность в связи с реализуемым в настоящее время по инициативе Президента РФ национальным проектом «Здоровье». В Послании Федеральному собранию в 2005 году В.В.Путин отметил, что **«необходимо возродить профилактику заболеваний как традицию российской медицинской школы»[12].** Иными словами, подтверждено, что предупреждение болезней и укрепление здоровье – самое главное, что должно занимать властные структуры, медицинскую науку и практическое здравоохранение. В связи с этим разрабатываемая система донозологического контроля физического здоровья и работоспособности с использованием показателей, хорошо понятных рядовым гражданам, может сыграть важную роль в перестройке общественного мировоззрения и формирования высокой культуры здоровья населения. Особую важность эта проблема приобретает в связи с поставленной 10 мая 2006 года Президентом в послании Федеральному собранию РФ важнейшей задачей – повышению рождаемости. Ни для кого не секрет, что произвести здоровое потомство могут только физически здоровые женщины и мужчины, ведущие здоровый образ жизни.

Существует много тестов для физического здоровья, однако отсутствуют модели прогнозирования в мире. На пример, в 2001 году Campbell PT, Katzmarzyk PT, Malina RM, Rao DC, Pérusse L, Bouchard C. разработаны модели прогнозирования физической активности и физической работоспособности (PWC150) в молодом возрасте с детства и подросткового возраста с учетом родительской меры в Норт-Йорке, Онтарио, Канада [13]. А также в 2003 году Trudeau F, Shephard RJ, Arsenault F, Laurencelle L. разработаны регрессионные модели для отслеживания физической подготовки с детства к взрослой жизни в Квебеке [14].

Целью работы является разработка статистических моделей для прогнозирования физического здоровья с использованием метода множественного регрессионного анализа для подростков в возрасте от 8 до 17 лет.

1. Метод анализа

В данной статье, будет представлена только модель для пятнадцатилетних девушек. Приведенный анализ 480 девушек в возрасте от 14 до 17 лет в медицинской компании «Народный Спорт Парк» [11] применен в данной работе. Выборка по каждой группе составляла 120 человек.

На основе полученных результатов анализа прогнозируем физическую работоспособность человека (PWC170/кг), которая является одной из важнейших компонентов физического здоровья человека, характеризующей способность организма эффективно выполнять большую мышечную работу и противостоять утомлению. Уровень общей выносливости определяется возможностями мышечной, дыхательной, сердечнососудистой, нервной, эндокринной систем, слаженность их работы при физических нагрузках и, в конечном счете, может служить обобщенной оценкой физического состояния организма.

После измерения морфологических и функциональных показателей физического здоровья человека проведен анализ полученных результатов. Предполагаем, что построить прогноз значений параметра PWC (физическая работоспособность) с помощью множественной регрессии. В этом случае необходимо выяснить математическую зависимость физической работоспособности человека от измеряемых морфологических и функциональных показателей.

В данном случае, для анализа используются двенадцать морфологических и функциональных показателей для группы девушек и мальчиков в возрасте от 14 до 17 лет, представленных в таблице 1.

Таблица 1

Параметры для регрессионного анализа

Символ	Определение символа	Сокращение Определения символов
Y	отклик (PWC170- физическая работоспособность) кгм/кг в мин	PWC170/кг
X_1	жизненная емкость легких, мл	ЖЕЛ
X_2	пульс в покое (частота сердечных сокращений, уд/мин)	ЧСС
X_3	систолическое артериальное давление, мм.рт.ст.	АД-С
X_4	диастолическое артериальное давление, мм.рт.ст.	АД-Д
X_5	задержка дыхания, сек	Гипокс.
X_6	весоростовой коэффициент (Кетле), гр/см	Кетле
X_7	гибкость позвоночника, см	Гибк.
X_8	координация движения (бросание в стену теннисных мячей, количество пойманных мячей из 6)	Коорд
X_9	зрительно-двигательная реакция (тест с падающей линейкой, см)	ЗРД
X_{10}	мышцы плечевого пояса (отжимание)	Отжим
X_{11}	мышцы брюшного пресса (пресс)	Пресс
X_{12}	тест Руфье (приседание)	Руфье

В работе используется множественный регрессионный анализ для оценки физической работоспособности. Регрессионный анализ является методом моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель имеет функцию независимой переменной и параметров с добавленной случайной переменной. Множественная регрессионная модель населения представляет в следующем виде:

$$Y_i = B_0 + B_1 X_1 + B_2 X_2 + ... + B_n X_n + \varepsilon, \quad (1)$$

где Y - отклик (зависимая переменная), B_0 - оценка постоянной составляющей, B_i – *i*-ый коэффициент множественной регрессии, X_i – *i*-ая независимая переменная, ε – ошибка; ($i=0,1,..,n$).

В матричной форме множественная регрессионная модель имеет вид:

$$Y = XB + \varepsilon, \quad (2)$$

где Y — вектор столбец наблюдений, размерность m×1;

X — матрица независимых переменных, размерность m×n;

B — вектор столбец параметров, подлежащих оцениванию, размерность n×1 (коэффициентов регрессии);

ε — случайный вектор-столбец размерности *n* x *1* ошибок наблюдений (остатков).

Оборудованная множественная регрессионная модель имеет вид:

$$\hat{y}_i = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n, \quad (3)$$

в матричном виде:

$$\hat{Y} = X\hat{\beta}. \quad (4)$$

Для получения коэффициентов регрессии нужно использовать метод наименьших квадратов, Метод основан на минимизации суммы квадратов остатков регрессии. Согласно методу наименьших квадратов, вектор-столбец оценок коэффициентов регрессии, получается, по формуле:

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad (5)$$

где X^T – транспонированная матрица независимых переменных, $(X^T X)^{-1}$ – обратная матрица, Y – вектор измерений.

2. Регрессионный анализ для пятнадцатилетних девушек

Запускаем регрессионный анализ (использовано программное обеспечение MS Excel и SPSS) и рассмотрим приведенные результаты анализа для 120-ти пятнадцатилетних девушек. В этом случае после вычисления результатов регрессионного анализа, в первую очередь, нужно проверить, что полученная модель является статистически значимой. Для этого можно использовать метод дисперсионного анализа. Напомним, что дисперсионный анализ позволяет выявить зависимости в экспериментальных данных путём исследования значимости различий в средних значениях.

В соответствии с дисперсионного анализа, выяснить полезность линии регрессии можно с помощью величины F – статистики. F - статистика является отношением объясненной дисперсии (среднеквадратическая регрессия) и необъясненной дисперсии (среднеквадратическая ошибка). В этом случае объясненную дисперсию или среднеквадратическую регрессию (MSR - Mean Squared Regression) можно вычислить по формуле [1-7]

$$MSR = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{k}. \quad (6)$$

Необъясненная дисперсия или среднеквадратическая ошибка (MSE- Mean Squared Error) вычисляет по формуле [1-7]

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-k-1}. \quad (7)$$

Здесь MSR – среднеквадратическая регрессия (объясненная дисперсия), MSE – среднеквадратическая ошибка (необъясненная дисперсия), y_i - i-ая зависимая переменная PWC, $\hat{y}_i$ – i-ое предсказанное значение параметра, $\bar{y}$ – среднее значение зависимой переменной, k – количество регрессоров, n – количество наблюдения; ($i=0,1,...,n$).

Величину F– статистику можно вычислить по формуле [2-7]

$$F = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \Big/ k}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \Big/ (n-k-1)} \text{ (или) } F = \frac{MSR}{MSE}. \quad (8)$$

Для проверки значимости модели будем использовать F таблицы распределения (F-распределения Фишера). В данном случае, использование F - критерия сводится к сравнению величины F – статистики со значением F – критерий в F – таблице. В случае если вычисленная величина F – статистики больше или равна критической величине F - критерий в F – таблице. Отсюда следует, что регрессионная модель становится статистически значимой. В таблице 2 приведены результаты дисперсионного анализа.

Таблица. 2

Дисперсионный анализ (ANOVA) для 15 – летних девушек

ANOVA					
	Степень свободы	сумма квадратов	Дисперсия на степень свободы (среднее квадратов)	F	Значимость F
Регрессия	12	246.4964473	20.5413706	4.7144844	0.0000039
Ошибка	107	466.2072194	4.3570768		
Общий	119	712.7036667			

Таким образом, для множественной регрессионной модели пятнадцатилетних девушек вычисленная F– статистика равна 4,7168, и из F-таблицы распределения F – критерия равна 1,8337 [10]. Соответственно величина F – статистики больше чем, величина F – критерия в F – таблице распределения, и что регрессионная модель является статистической значимой и полезной.

Попробуем количественно оценить, насколько полезна полученная множественная регрессионная модель для прогнозирования параметра PWC, другими словами, какую часть в величине прогноза параметра PWC обосновывает множественная регрессионная модель.

Для указанного количественного оценивания полезности полученной множественной регрессионной модели была исследована попытка воспользоваться коэффициентом детерминации R2 [2-7]. Для того чтобы была возможность сравнивать модели с разным числом факторов, так чтобы число факторов (регрессоров) не влияло на значение R^2, обычно используется *скорректированный коэффициент детерминации*. Коэффициент детерминации R2 и скорректированный коэффициент детерминации $R2_a$ можно вычислить по формулам

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (9)$$

$$R^2{}_a = 1 - \left(\left(\frac{n-1}{n-k-1}\right)(1-R^2)\right), \quad (10)$$

где y_i - i-ая зависимая переменная PWC, $\hat{y}_i$ – i-ое предсказанное значение параметра, $\bar{y}$ – среднее значение зависимой переменной, k – количество регрессоров, n – количество наблюдения; ($i=0,1,\ldots,n$). Для пятнадцатилетних девушек, скорректированного коэффициента

детерминации $R2_a$ равно 0,27 (или 27%). Таким образом, в величине прогноза параметра PWC многофакторная модель обосновывает 27% дисперсии.

Оценим качество модели с помощью стандартной ошибки оценки S_ε, вычисляемой по формуле [2-7]

$$S_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}}. \quad (11)$$

Для пятнадцатилетних девушек, вычисленное значение стандартной ошибки оценки S_ε равно 2,087. В таблице 3 приведены выводы регрессионного анализа.

Таблица 3

Регрессионный вывод для 15 летних девушек

Модель	R	R2	$R2_a$	Стандартная ошибка оценки
15 девочек	0,588	0,346	0,272	2.08736

В таблице 4 иллюстрированы вычисленные результаты коэффициентов регрессии, величины т-статистики и т.д. По таблице можем создать регрессионную модель, используя соотношение (3). В этом случае проверяются вычисленные величины коэффициентов регрессии, которые являются статистическими значимыми. Для ответа на поставленный вопрос использовались следующие показатели: стандартная ошибка каждого из коэффициентов регрессии, т – статистика и Р – значение (P-value), которые очень важны для проверки гипотез значимости каждого из коэффициентов. Для проверки гипотез на самом деле нужно знать значение т – статистики каждого коэффициента, поскольку т – статистика позволяет проверить значимости каждого из коэффициентов регрессии [1-7]. В частности, т – статистику для каждого коэффициентов можно вычислить с помощью стандартной ошибки каждого из коэффициента регрессии. Следовательно, в первую очередь нужно вычислить стандартные ошибки коэффициентов регрессии. Для вычисления стандартной ошибки коэффициентов b_0, b_1, …, b_{12} используется формула [1-7]

$$S(b) = \sqrt{S_{ост}^2 . (X^T X)^{-1}}, \quad (12)$$

где

$$S_{ост}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}. \quad (13)$$

Здесь *S(b)* – стандартная ошибка регрессионных коэффициентов, $(X^T X)^{-1}$ – обратная матрица, y_i – измеряемые значения PWC (зависимая переменная), $\hat{y}$ – *i*-ое предсказанное значение параметра PWC, (*i*=1,…,*n*), *k* – степень свободы регрессии или число регрессоров, *n* – количество наблюдений.

Затем можно вычислить т-статистику для каждого коэффициента по формуле [1-6]

$$t(b) = \frac{b}{S_b}, \tag{14}$$

где *t(b)* – т - статистика одного коэффициента, *b* - коэффициент регрессии, S_b - стандартная ошибка коэффициента регрессии.

Для проверки гипотез значимости коэффициента будем использовать критерий Стьюдента. В нашем случае использование критерия Стьюдента (Т-критерия) сводится к сравнению значении т-статистики со значение Т-критерий в Т-таблице [2-8]. В случае, если вычисленное т-значение больше или равно критическому значению Т в Т-таблице, можно сделать вывод, что регрессионный коэффициент является статистическим значимым.

В таблице 4 значения т-статистики ($t_{b2,b6,b8,b9,b11,b12}$) по сравнению со значением Т-критерия ($t_{табл}$ = 1,646 (n=120, α=0,05%)) получаются большими[9]. Соответственно, эти коэффициенты (b_2, b_6, b_8, b_9, b_{11} и b_{12}) т -статистики являются статистическими значимыми.

Также рассмотрена проверка мультиколлинеарности в таблице 4. В этом случае мультиколлинеарность представляет собой статистический феномен, в котором два или более регрессоров сильно коррелируют в модели множественной регрессии. Проведен анализ мультиколлинеарности с помощью метода VIF (Variance Inflation Factor). При VIF больше пяти мультиколлинеарность существует в модели и она не является правильной. Все величины VIF меньше пяти. Поэтому мультиколлинеарность отсутствует в модели. Для вычисления VIF можно использовать формулу [7]

$$VIF_j = \frac{S_{x_j}^2 (n-1) S_b{}^2}{S_{ост}^2}, \tag{15}$$

где VIF - отклонение коэффициента инфляции (Variance Inflation Factor), S_{x_j} - стандартное отклонение xj, S_{b_j} - стандартная ошибка коэффициента регрессии, $S_{ост}^2$ - среднеквадратическая остаточная.

Для группы из 120 девушек пятнадцать лет были вычислены коэффициенты b_0 .. b_{12} с использованием формулой (5). В таблице 4 приведены вычисленные коэффициенты, стандартные ошибки коэффициентов регрессии, т-статистики, P-value и коллинеарность для пятнадцатилетних девушек.

Таблица 4

Коэффициенты регрессионного анализа, стандартные ошибки коэффициентов регрессии, т-статистики, P-value и коллинеарность для 15 летних девушек

	коэффициенты	**стандартные ошибки**	**т-статистики**	***P-value***	*Коллинеарность статистика*	
					толерантность	*VIF*
Константа	b_0=18,6596	2,7881	6,6926	0,0000		
ЖЕЛ	b_1=0,0003	0,0004	0,7374	0,4625	0,816	1,225
ЧСС	b_2=-0,0305	0,0169	-1,8073	0,0735	0,888	1,126
АД-С	b_3=-0,0135	0,0203	-0,6623	0,5092	0,551	1,813
АД-Д	b_4=-0,0073	0,0310	-0,2346	0,8150	0,504	1,983
гипокс	b_5=0,0097	0,0110	0,8793	0,3812	0,889	1,124
Кетле	b_6=-0,0091	0,0043	-2,1235	0,0360	0,788	1,269
гибк	b_7=0,0233	0,0274	0,8512	0,3966	0,892	1,121
коорд	b_8=-0,0474	0,0272	-1,7402	0,0847	0,917	1,090
ЗРД	b_9=0,0564	0,0257	2,1937	0,0304	0,871	1,149
отжим	b_{10}=-0,0072	0,0244	-0,2939	0,7694	0,824	1,214
пресс	b_{11}=0,0626	0,0354	1,7695	0,0797	0,853	1,172
Руфье	b_{12}=-0,4089	0,0858	-4,7652	0,0000	0,905	1,105

Уравнение для предсказания значения параметра PWC по таблице 4 и формуле (3) имеет вид:

$$\hat{y} = 18{,}6596 + 0{,}0003x_1 - 0{,}0305x_2 - 0{,}0135x_3 - 0{,}0073x_4 + 0{,}0097x_5 - 0{,}0091x_6 + 0{,}0233x_7 - 0{,}0474x_8 + 0{,}0564x_9 - 0{,}0072x_{10} + 0{,}0626x_{11} - 0{,}4089x_{12}, \quad (16)$$

где $\hat{y}$ – предсказанное значение параметра PWC, x_i – независимые переменные (таблица 1).

После получения регрессионной модели, в соответствии с тестированием т - статистики, пренебрегаем незначимые параметры. Поэтому без использования незначимых параметров перезапускаем регрессию. Считаем, что модель со всеми предикторами - полная модель. А модель, которая содержит лишь некоторые из этих предсказателей, называется уменьшенной моделью. После перезапуска регрессии рассматривается значимость модели.

В таблице 5 видно, что F – статистика равна 9,157 и сравним с F критерий в F таблице распределения. Значение F-статистики (9,157) больше чем, значение $F_{табл.}$ (2,175). Соответственно, будем считать, что можно отвергнуть нулевую гипотезу. Уменьшенная регрессионная модель является статистической значимой.

Таблица 5

Дисперсионный анализ для 15 летних девочек

ANOVA					
	Степень свободы	сумма квадратов	Дисперсия на степень свободы (среднее квадратов)	F	Значимость F
Регрессия	6	233.167	38.861	9.157	0.000
Ошибка	113	479.537	4.244		
Общий	119	712.704			

Кроме того, для изменчивости модели рассмотрим вычисленное значение скорректированного коэффициента детерминации R^2_a. В таблице 6 показано, что значение R^2_a равно 0,291 (29,1%), уравнение регрессии составляет 29,1% дисперсии результативного признака.

Таблица 6

Регрессионный вывод для 15 летних девочек

Модель	R	R2	$R2_a$	Стандартная ошибка оценки
15 девушек	0,572	0,327	0,291	2.06002

В таблице 7 представлены вычисленные регрессионные коэффициенты для уменьшенной модели, с помощью которых получается регрессионная модель (17), используя соотношение (3). После этого выполняется проверка гипотезы для коэффициента регрессии. В таблице 8 т - статистики всех коэффициентов b_i больше $t_{табл}$ = 1,645 (n=120, α=0,05%). Следовательно, все коэффициенты являются статистическими значимыми. Проверяются мультиколлинеарности для уменьшенной модели (таб.8), все величины VIF меньше чем пять. Соответственно, мультиколлинеарность отсутствует в модели.

Таблица 7

Коэффициенты регрессионного анализа, стандартные ошибки коэффициентов регрессии, т-статистики, P-value и коллинеарность для 15 летних девочек

	коэффициенты	**стандартные ошибки**	**т-статистики**	***P-value***	*Коллинеарность статистика*	
					толерантность	*VIF*
Констант	b_0=18,6021	2,2141	8,4018	0,0000		
ЧСС	b_1=-0,0347	0,0160	-2,1705	0,0321	0,965	1,037
Кетле	b_2=-0,0096	0,0039	-2,4603	0,0154	0,942	1,061
коорд	b_3=-0,0516	0,0262	-1,9717	0,0511	0,966	1,035
ЗРД	b_4=0,0584	0,0243	2,4002	0,0180	0,946	1,057
пресс	b_5=0,0614	0,0332	1,8515	0,0667	0,946	1,057
Руфье	b_6=-0,4216	0,0808	-5,2179	0,0000	0,994	1,006

Уравнение для предсказания значения параметра PWC согласно таблице 7 и по формуле(3) имеет вид:

$$\hat{y} = 18{,}6021 - 0{,}0347x_1 - 0{,}0096x_2 - 0{,}0516x_3 + 0{,}0584x_4 + 0{,}0614x_5 - 0{,}4216x_5, \quad (17)$$

где $\hat{y}$ – предсказанное значение параметра PWC, x_i – независимые переменные (таблица 1).

По результатам сравнения полной модели (16) с уменьшенной моделью (17) оказывается, что уменьшенная модель лучше полной модели. Для сравнения двух моделей с помощью формулы (теста Multiple Partial F):

$$F = \frac{(SSR_{полная} - SSR_{уменьшенная})/q}{MSE_{полная}}. \quad (18)$$

Здесь SSR – сумма квадратов регрессии, MSE – среднеквадратическая остаточная, q – разница между количествами регрессоров из двух модели.

3. Анализ остатков

Проанализированы остатки уменьшенной модели. Для проверки анализа остаток рассматривается стандартизированная остаточная гистограмма, в которой показаны остатки нормального распределения и общая форма является приемлемой. Применяя результаты тестов по математическому подходу Колмогорова-Смирнова и Shapiro-Wilk (таб.8) то есть величины значимости больше чем 0,05, починяются остатки по нормальному распределению. На рис. 1 приведены результаты проверки нормальности остатков.

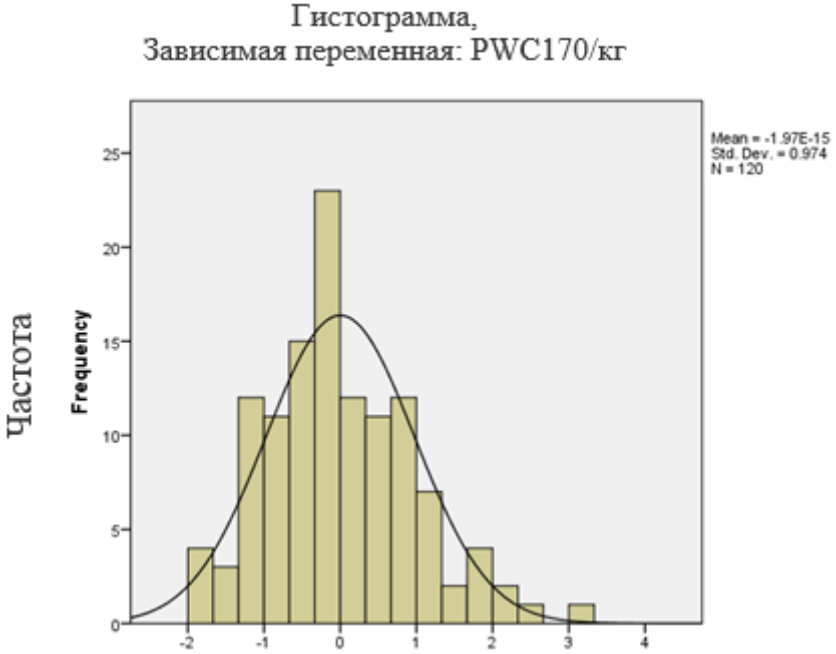

Рис. 1. *Гистограмма стандартизированных остатков*

Таблица 8

Результаты тестов Колмогорова Смирнова и Shapiro-Wilk.

Проверка нормальности остаток по математическому подходу						
	Колмогоров-Смирнов			Shapiro-Wilk		
	Статистика	df	Значи-мость.	Статистика	df	Значи-мость
Стьюдентизированные остаточные	0,071	120	0,200	0,982	120	0,112
Стандартизированные остаточные	0,072	120	0,196	0,981	120	0,091

Рассмотрим провертку гетероскедастичность модели. Под гетероскедастичностью понимают неравные дисперсии остатков модели (проблема в дисперсии остатков). В противном случае гетероскедастичность отсутствует гомоскедастичность, т.е равные дисперсии остатков. Можно проверить гетероскедастичность с использованием теста Бройша Пагана, теста Голдфелда, Уайта, Коэнкера и.т.д. Рассматривая точечную диаграмму (рис 3) для проверки гетероскедастичности остатков, в этой остаточной точечной диаграммы отсутствует шаблон. После вычисления, сравнивая величину Бройша Пагана с критерием хи-квадрат в таблице распределения хи-квадрата, получим величину Бройша Пагана = 10,498 и в таблице критическая величина хи-квадрата = 12,592.

Очевидно, что величина Бройша Пагана меньше критерия хи-квадрат, что означает, отсутствует гетероскедастичность в уменьшенной регрессионной модели. На рис 2 приведена точечная диаграмма стандартизированных остатков для проверки гетероскедастичности.

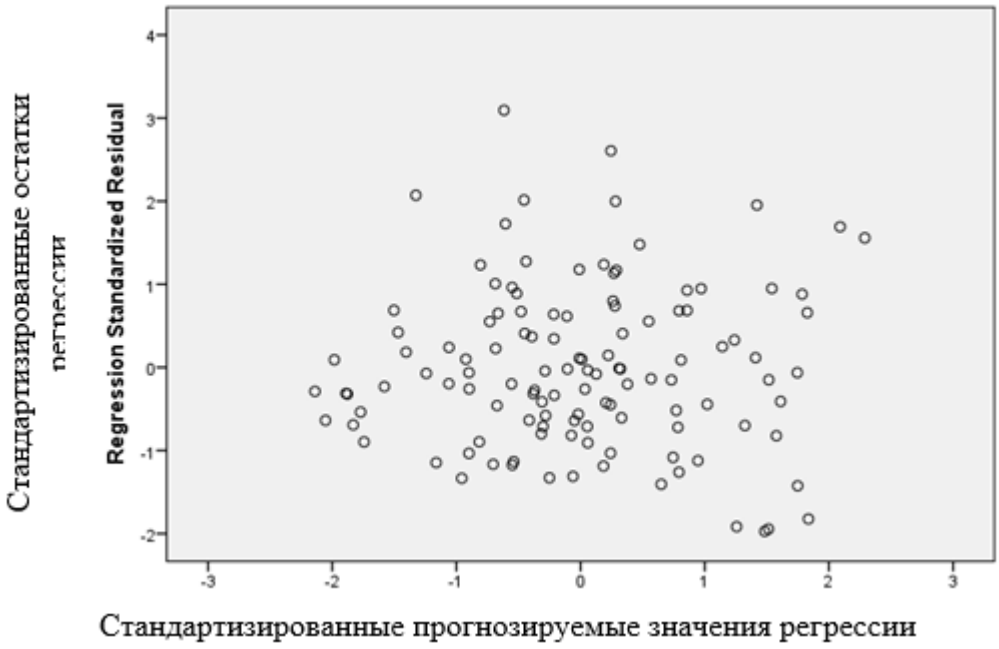

Рис. 2. *Точечная диаграмма остатков для15 летних девочек*

Введем автокорреляцию, т.е статистическую взаимосвязь между последовательностями величин одного ряда, взятых со сдвигом, например, для случайного процесса со сдвигом по времени. В этом случае, используем тест Дарбина Уотсона для обнаружения автокорреляции. Если остаточный e_t связан с наблюдением в момент времени T, можно записать тест статистики вида

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t-1}^{T} e_t^2}, \quad (19)$$

где Т-число наблюдений, e_t – остатки регрессионной модели.

Для проверки положительной автокорреляции на значении α тестовая статистика d по сравнению с нижним и верхним критическим значением из таблицы Дарбина Уотсона ($d_{L,\alpha}$ and $d_{U,\alpha}$) должна удовлетворять следующим условиям:

- если $d < d_L$, при α=0,05, то имеется статистические доказательства, что в остатках есть положительная автокорреляция;
- если $d > d_U$, при α=0,05, то отсутствуют статистические доказательства того, что в остатках нет автокорреляции;
- если d_L, $\alpha < d < d_U$, при α=0,05, то тест является не решенным (тест считается не прошедшим).

Таким образом, по тесту Дарбина Уотсона d = 2,238 и в таблице указывается $d_{\text{нижняя}}$ =1,651, $d_{\text{верхняя}}$ = 1,817 для шести степеней свободы. Поэтому величина Дарбина Уотсона $d > d_{\text{верхняя}}$ и можно считать, что положительная автокорреляция отсутствует в модели.

В результате получим, что после проверки гипотезы сформулированной в данной статье (F-тест, t-тест, и.т.д) модель множественной регрессии физической работоспособности пятнадцатилетних девушек является полезной и найдено применение для оценки состояния здоровья подростков. Даже независимые и зависимые переменные имеют слабую корреляцию. Другие регрессионные модели для 14, 15 и 17 летних девушек также позволяют получить аналогичные результаты.

Заключение

Для прогнозирования физического здоровья человека выбран метод множественного регрессионного анализа статистики, который позволяет проводить анализ многофакторных статистических моделей. Разработаны математические и статистические модели для прогнозирования физического здоровья девушек в возрасте от четырнадцати до семнадцати лет. Определены значимые параметры для математических моделей прогнозирования, с помощью которых быстро и эффективно можно оценить физическое здоровье девушек в возрасте от четырнадцати до семнадцати лет. По результатам регрессионного анализа регрессионные модели девушек в возрасте от четырнадцати до семнадцати лет являются статистически значимыми и могут быть использованы при оценке состояния здоровья. Найдены значимые параметры для оценки физической работоспособности прогнозирования физического здоровья девушек в возрасте от четырнадцати до семнадцати лет.

ЛИТЕРАТУРА

1. Математическая статистика: Учеб. Для вузов /В.Б. Горянинов, И.В. Павлов, Г.М. Крищенко. -2-е изд., стереотип. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2002. – 424 с. (Сер. Математика в техническом университете; Вы. XVII).
2. Applied regression analysis: a research tool. - 2nd ed. / John O. Rawlings, Sastry G. Pentula, David A. Dickey. Изд-во - (Springer texts in statistics), - 671p.
3. Statistical Models: Theory and Practice, David A. Freedman, Cambridge University Press (2005), - 414p.
4. Modeling and interpreting interactive hypotheses in regression analysis, Cindy D. Kam and Robert J. Franzitsi Jr., University of Michigan Press (2009), - 168p.
5. Regression analysis by example. - 4th ed. / Samprit Chatterjee, Ah S. Hadi, Wiley series in probability and statistics Established by Walter A. Shewhart and Samuel S. Wilks, - 366p.
6. Multiple regression in behavioral research (Explanation and prediction), - 3rd ed / Elazar J. Pedhazur. Изд-во - Thomson Learning, 1997. – 1072p.
7. Linear Regression Analysis: Assumptions and Applications, John P. Hoffmann Department of Sociology Brigham Young University (2005), - 259p.
8. Six sigma online [Электронный ресурс]. Режим доступа: http://sixsigmaonline.ru/load/24-1-0-210 (дата обращения 20.11.2013).
9. Google document T - таблица [Электронный ресурс]. Режим доступа: https://docs.google.com/viewer?a=v&q=cache:1boQad1pHCQJ:www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf (дата обращения 20.11.2013).
10. Statistics Online Computational Resource (SOCR) [Электронный ресурс]. Режим доступа: http://socr.ucla.edu/Applets.dir/F_Table.html (дата обращения 20.11.2013)
11. Народный СпортПарк [Электронный ресурс]. Режим доступа: http://sportpark.ru/ (дата обращения 20.11.2013).
12. Address to Federal Assembly Russia [Электронный ресурс]. Режим доступа: http://archive.kremlin.ru/eng/speeches/2005/04/25/2031_type70029type82912_87086.shtml (дата обращения 11.05.2013).
13. Campbell PT, Katzmarzyk PT, Malina RM, Rao DC, Pérusse L, Bouchard C. Prediction of physical activity and physical work capacity (PWC150) in young adulthood from childhood and adolescence with consideration of parental measures. Abstract, 2001. [Электронный ресурс]. Режим доступа: http://www.ncbi.nlm.nih.gov/pubmed/11460863 (дата обращения 11.05.2013).
14. Trudeau F, Shephard RJ, Arsenault F, Laurencelle L. Tracking of physical fitness from childhood to adulthood. Abstract, 2003. [Электронный ресурс]. Режим доступа: http://www.ncbi.nlm.nih.gov/pubmed/12825334?report=abstract (дата обращения 11.05.2013).

Рецензент: Симаранов. С. Ю., Генеральный директор ЗАО « Техноконсалт», доктор технических наук, профессор.

Kyi Thar Soe
Bauman Moscow State Technical University "BMSTU"
Russia, Moscow
E-Mail: kyithar82@gmail.com

Development of mathematical models and software for the human's physical health

Abstract. The paper presents results of research for the mathematical model of the physical health of a "healthy person" in the regression analysis, where the factors are the physical parameters of the person, and as a response - an indicator of physical working capacity. The regression equations were constructed for each age group. After performing multiple regressions analysis obtained multiple regression models that can predict the physical working capacity for girls aged fourteen to seventeen years. Article shows the results of the analysis and development of the regression models and program software for the girls of fifteen. Also presents the hypotheses testing, ie checking the significance of the model, the significance of the coefficients, heteroscedasticity, autocorrelation, multicollinearity and normality. In conclusion, for the prediction of physical health method selected multiple regression statistics, which allows the analysis of multivariate statistical models. Defined the relevant parameters for the mathematical prediction models, with which you can quickly and efficiently assess the physical health of girls between the ages of fourteen to seventeen years.

Keywords: regression model; correlation; analysis of variance; t - statistics; F-statistics; the coefficient of determination; heteroscedasticity.

Identification number of article 37TVN314

REFERENCES

1. Matematicheskaja statistika: Ucheb. Dlja vuzov /V.B. Gorjaninov, I.V. Pavlov, G.M. Krishhenko. -2-e izd., stereotip. – M.: Izd-vo MGTU im. N.Je. Baumana, 2002. – 424 s. (Ser. Matematika v tehnicheskom universitete; Vy. XVII).
2. Applied regression analysis: a research tool. - 2nd ed. / John O. Rawlings, Sastry G. Pentula, David A. Dickey. Izd-vo - (Springer texts in statistics), - 671p.
3. Statistical Models: Theory and Practice, David A. Freedman, Cambridge University Press (2005), - 414p.
4. Modeling and interpreting interactive hypotheses in regression analysis, Cindy D. Kam and Robert J. Franzitsi Jr., University of Michigan Press (2009), - 168p.
5. Regression analysis by example. - 4th ed. / Samprit Chatterjee, Ah S. Hadi, Wiley series in probability and statistics Established by Walter A. Shewhart and Samuel S. Wilks, - 366p.
6. Multiple regression in behavioral research (Explanation and prediction), - 3rd ed / Elazar J. Pedhazur. Izd-vo - Thomson Learning, 1997. – 1072p.
7. Linear Regression Analysis: Assumptions and Applications, John P. Hoffmann Department of Sociology Brigham Young University (2005), - 259p.
8. Six sigma online [Jelektronnyj resurs]. Rezhim dostupa: http://sixsigmaonline.ru/load/24-1-0-210 (data obrashhenija 20.11.2013).
9. Google document T - tablica [Jelektronnyj resurs]. Rezhim dostupa: https://docs.google.com/viewer?a=v&q=cache:1boQad1pHCQJ:www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf (data obrashhenija 20.11.2013).
10. Statistics Online Computational Resource (SOCR) [Jelektronnyj resurs]. Rezhim dostupa: http://socr.ucla.edu/Applets.dir/F_Table.html (data obrashhenija 20.11.2013)
11. Narodnyj SportPark [Jelektronnyj resurs]. Rezhim dostupa: http://sportpark.ru/ (data obrashhenija 20.11.2013).
12. Address to Federal Assembly Russia [Jelektronnyj resurs]. Rezhim dostupa: http://archive.kremlin.ru/eng/speeches/2005/04/25/2031_type70029type82912_87086.shtml (data obrashhenija 11.05.2013).
13. Campbell PT, Katzmarzyk PT, Malina RM, Rao DC, Pérusse L, Bouchard C. Prediction of physical activity and physical work capacity (PWC150) in young adulthood from childhood and adolescence with consideration of parental measures. Abstract, 2001. [Jelektronnyj resurs]. Rezhim dostupa: http://www.ncbi.nlm.nih.gov/pubmed/11460863 (data obrashhenija 11.05.2013).
14. Trudeau F, Shephard RJ, Arsenault F, Laurencelle L. Tracking of physical fitness from childhood to adulthood. Abstract, 2003. [Jelektronnyj resurs]. Rezhim dostupa: http://www.ncbi.nlm.nih.gov/pubmed/12825334?report=abstract (data obrashhenija 11.05.2013).