

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 8, №1 (2016) <http://naukovedenie.ru/index.php?p=vol8-1>

URL статьи: <http://naukovedenie.ru/PDF/53TVN116.pdf>

DOI: 10.15862/53TVN116 (<http://dx.doi.org/10.15862/53TVN116>)

Статья опубликована 04.03.2016.

Ссылка для цитирования этой статьи:

Куи Тар Со Разработка программного обеспечения для прогнозирования физической работоспособности человека на основе метода регрессионного анализа // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 8, №1 (2016) <http://naukovedenie.ru/PDF/53TVN116.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI: 10.15862/53TVN116

УДК 004.4, 519.2

Куи Тар Со

ФГБОУ ВПО «Московский государственный технический университет имени Н.Э. Баумана», Россия, Москва¹
Аспирант

E-mail: kyithar82@gmail.com

Разработка программного обеспечения для прогнозирования физической работоспособности человека на основе метода регрессионного анализа

Аннотация. В статье представлены результаты исследований по созданию статистической модели физической работоспособности «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 получены множественные регрессионные модели, которые могут предсказать физическую работоспособность для подростков в возрасте от восьми до семнадцати лет. В статье приведены результаты анализа и разработка регрессионной модели для семнадцатилетних мальчиков. Также представлены проверки гипотез для модели, т.е. проверка значимости модели, значимости коэффициентов, гетероскедастичности, автокорреляции, мультиколлинеарности и нормальности остатков. В результате для прогнозирования физической работоспособности человека выбран метод множественного регрессионного анализа статистики, который позволяет проводить анализ многофакторных статистических моделей. Разработаны регрессионные модели и программное обеспечение для прогнозирования физической работоспособности человека в возрасте от восьми до семнадцати лет. Определены значимые параметры для статистических моделей прогнозирования, с помощью которых быстро и эффективно можно оценить физическое здоровье человека в возрасте от восьми до семнадцати лет.

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

¹ 105005, Россия, Москва, Госпитальный пер., д. 4/6, корп. 1

Введение

В настоящее время мало кто уделяет внимание физкультуре. Все больше и больше времени люди проводят перед компьютером и телевизором. От этого мышцы спины, ног, шеи и т.д. слабеют, а человек больше подвержен болезням. С другой стороны, 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»** [9]. 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 донологического 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. Для этих ситуации физическая работоспособность человека станет важной для здоровья в человечестве.

Физическая работоспособность - один из важнейших компонентов физического здоровья человека. Она характеризует способность организма эффективно выполнять большую мышечную работу и противостоять утомлению. Уровень общей физической работоспособности определяется возможностями мышечной, дыхательной, сердечнососудистой, нервной, эндокринной систем и слаженностью их работы при физических нагрузках. Физическая работоспособность рассчитывается в величинах мощности выполняемой нагрузки (кгм/мин) в расчете на 1 кг массы тела, соответствующей ЧСС 170 уд/мин.

Целью работы является разработка программного обеспечения статистических моделей для прогнозирования физической работоспособности с использованием метода множественного регрессионного анализа для подростков в возрасте от 8 до 17 лет.

1. Метод анализа и параметры

Для прогнозирования физической работоспособности используем метод множественного регрессионного анализа. 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:

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

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; ($i=0, 1, \dots, n$).

В матричной форме множественная регрессионная модель имеет вид:

$$Y = XB + \varepsilon,$$

где Y — вектор столбец наблюдений, размерность $m \times 1$;

X — матрица независимых переменных, размерность $m \times n$;

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

ε — случайный вектор-столбец размерности $n \times 1$ ошибок наблюдений (остатков).

В анализе используются двенадцать морфологических и функциональных показателей для группы девушек и мальчиков в возрасте от 8 до 17 лет, представленных в таблице 1. 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.

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 данной статье, будет представлена только модель для семнадцатилетних мальчиков. Приведенный анализ 2400 мальчиков и девушек в возрасте от 8 до 17 лет в медицинской компании «Народный Спорт Парк» [8] применен в данной работе. Выборка по каждой группе составляла 120 человек.

2. Регрессионный анализ для семнадцатилетних мальчиков

Запускаем регрессионный анализ (использовано программное обеспечение MS Excel и SPSS) и рассмотрим приведенные результаты анализа для 120-ти пятнадцатилетних девушек. После вычисления регрессии с помощью программных пакетов, в первую очередь, необходимо смотреть в таблицу дисперсионного анализа. Для проверки значимости уравнения или качества уравнения регрессии в целом производится на основе **F-критерия**

Фишера, которому предшествует дисперсионный анализ. Дисперсионный анализ можно вычислить по таблице 2.

Таблица 2

Схема дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на степень свободы (среднее квадратов)
Факторная (регрессия)	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	k	$S_{факт}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{k}$
Остаточная	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n-k-1$	$S_{ост}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}$
Общая	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n-1$	$S_{общ}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$

Здесь n – число наблюдений, k – число параметров при переменной i , поэтому F -критерия Фишера можно вычислить по формуле:

$$F = \frac{S_{факт}^2}{S_{ост}^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / k}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-k-1)}$$

Фактическое значение F -критерия Фишера сравнивается с табличным значением $F_{табл.}$ (α, k_1, k_2) при заданном уровне значимости α и степенях свободы $k_1 = k$ и $k_2 = n-k-1$. При этом, если фактическое значение F -критерия больше табличного $F_{факт} > F_{табл.}$, то признается статистическая значимость уравнения в целом. В соответствии с дисперсионным анализом, выяснить полезность линии регрессии можно с помощью величины F -статистики. В этом случае, вычисленная F -статистика = 3,0298 (и сравним F -критерия в F таблице = 1,8337), вычисленная F -статистика больше F -критерия и можно сказать, что регрессионная модель является статистически значимым. В таблице 3 приведены вычисленные результаты дисперсионного анализа для шестнадцатилетних мальчиков.

Таблица 3

Дисперсионный анализ (ANOVA) для семнадцатилетних мальчиков

ANOVA	Степень свободы	сумма квадратов	Дисперсия на степень свободы (среднее квадратов)	F	Значимость F
Факторная (Регрессии)	12	334,0213	27,8351	5,3844	0,0000
Остаточная	107	553,1454	5,1696		
Общая	119	887,1667			

После проверки значимости уравнения регрессии можно рассмотреть **коэффициента множественной детерминации** R^2 . Использование R^2 для оценки качества модели обладает тем недостатком, что включение в модель нового фактора (даже несущественного) автоматически увеличивает величину R^2 . Поэтому, при большом количестве факторов, предпочтительнее использовать, так называемый **улучшенный, скорректированный коэффициент множественной детерминации** R^2 , определяемый соотношением:

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

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

В таблице 4 приведены вычисленные результаты R^2 , R_a^2 и стандартной ошибки оценки. В этом случае, регрессионная модель означает, что 15,8% изменчивости зависимой переменной Y . Это объясняется изменчивостью независимых переменных X_i . Здесь стандартная ошибка оценки - это мера точности предсказаний. В линии регрессии, тем меньше стандартная ошибка оценки, тем точнее предсказания. стандартная ошибка оценки можно вычислить по формуле:

$$\sigma_{est} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}},$$

где σ_{est} - стандартная ошибка оценки, \hat{y} - предсказанное значение, y - отклик, n - число наблюдение.

Таблица 4

Регрессионный вывод для 10 летних девочек

Модель	R	R^2	R_a^2	Стандартная ошибка оценки
17 - лет мальчиков	0,614	0,377	0,307	2,274

После этого необходимо рассмотреть коэффициенты регрессионной модели. Для получения коэффициентов регрессии можно использовать метод наименьших квадратов, который основан на минимизации суммы квадратов остатков регрессии. Согласно методу наименьших квадратов, вектор-столбец оценок коэффициентов регрессии получается по формуле:

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

где X^T - транспонированная матрица независимых переменных, $(X^T X)^{-1}$ - обратная матрица от произведения транспонированной матрицы независимых переменных на саму матрицу независимых переменных, Y - вектор наблюдений.

Для проверки значимости коэффициентов регрессии необходимо использовать критерий Стьюдента. Использование критерия Стьюдента (Т-критерия) сводится к сравнению

значения т-статистики со значением Т-критерия в Т-таблице с учетом принятого уровня значимости (α) и числа степеней свободы вариации $k(v) = n-2$. В случае, если вычисленное т-значение больше или равно критическому значению Т в Т-таблице, можно сделать вывод, что регрессионный коэффициент является статистически значимым. т-статистика можно вычислить по формуле:

$$t(b) = \frac{b}{S_{b_i}},$$

где, $t(b)$ - т-статистика одного коэффициента, b - коэффициент регрессии i , S_{b_i} - стандартная ошибка коэффициента регрессии. Здесь стандартная ошибка коэффициента регрессии можно вычислить по формуле:

$$S_b = \sqrt{S_{ocm}^2 \cdot (X^T X)^{-1}}; S_{ocm}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}.$$

Далее проверяются мультиколлинеарности для модели. Мультиколлинеарность представляет собой статистический феномен, в котором два или более регрессора сильно коррелируют в модели множественной регрессии. Можно проверить с использованием VIF. Если VIF больше 5, то мультиколлинеарность существует в модели:

$$VIF_j = \frac{S_{x_j}^2 (n-1) S_{b_j}^2}{S_{ocm}^2}.$$

Здесь, VIF - отклонение коэффициента инфляции (Variance Inflation Factor), S_{x_j} - стандартное отклонение x_j , S_{b_j} - стандартная ошибка коэффициента регрессии, S_{ocm}^2 - средне-квадратическая остаточная. В таблице 5, приведены вычисленные значения коэффициентов регрессии, величины т-статистики, VIF и т.д. В таблице, видно что, т-статистики для коэффициентов (b_1, b_2 и b_{12}) больше чем Т-критерия ($t_{b_1, b_2, b_{12}} > t_{табл} = 1.646$ ($n=120, \alpha=0.05\%$)). Соответственно т-статистики, эти коэффициенты являются статистическими значимыми. Величины VIF меньше чем 5 и можно считать, что мультиколлинеарность отсутствует в модели.

Таблица 5

Коэффициенты регрессионной модели семнадцатилетних мальчиков

	коэффициенты	стандартные ошибки	т-статистики	P-value	Коллинеарность статистика	
					толерантность	VIF
Констант	$b_0 = 17,2791$	2,8492	6,0646	0,0000		
ЖЕЛ	$b_1 = 0,0006$	0,0003	2,2088	0,0293	0,843	1,186
ЧСС	$b_2 = -0,0359$	0,0157	-2,2898	0,0240	0,818	1,223
АД-С	$b_3 = 0,0053$	0,0206	0,2546	0,7996	0,601	1,663
АД-Д	$b_4 = -0,0144$	0,0249	-0,5791	0,5638	0,723	1,383
гипокс	$b_5 = 0,0095$	0,0092	1,0297	0,3055	0,908	1,102
Кетле	$b_6 = 0,0006$	0,0041	0,1361	0,8920	0,829	1,206
гибк	$b_7 = 0,0222$	0,0295	0,7522	0,4536	0,856	1,169
коорд	$b_8 = 0,0219$	0,0608	0,3604	0,7192	0,858	1,165
ЗРД	$b_9 = -0,0102$	0,0278	-0,3659	0,7152	0,934	1,071
отжим	$b_{10} = 0,0163$	0,0261	0,6264	0,5324	0,800	1,249
пресс	$b_{11} = 0,0330$	0,0371	0,8889	0,3760	0,879	1,138
Руфье	$b_{12} = -0,6429$	0,1015	-6,3363	0,0000	0,928	1,077

Для прогнозирования значения параметра $PWC_{170}/кг$ необходимо подставить в формулу (1) значения коэффициентов из таблицы 5:

$$\hat{y} = 17,2791 + 0,0006x_1 - 0,0359x_2 + 0,0053x_3 - 0,0144x_4 + 0,0095x_5 + 0,006x_6 + 0,0222x_7 + 0,0219x_8 - 0,0102x_9 + 0,0163x_{10} + 0,0330x_{11} - 0,6429x_{12}, \quad (2)$$

После получения регрессионной модели, в соответствии с тестированием t -статистики пренебрегаем незначимыми параметрами. 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 уменьшенной модели. В таблице 6 показано, что F -статистика равна 20,8144, сравним с F -критерий в F таблице распределения. Значение F -статистики (20,8144) больше, чем значение F -критерия ($F_{табл.} = 2,6802$). Соответственно, будем считать, что можно отвергнуть нулевую гипотезу. Поэтому, уменьшенная регрессионная модель является статистической значимой. Кроме того, для изменчивости модели рассмотрим вычисленное значение скорректированного множественного коэффициента детерминации R^2_a . В таблице 7 показано, что значение R^2_a равно 0,333 (33,3%), уравнение регрессии составляет 33,3% дисперсии результативного признака.

Таблица 6

Дисперсионный анализ для семнадцатилетних мальчиков

ANOVA	Степень свободы	сумма квадратов	Среднеквадратическая	F	Значимость F
Регрессия	3	310,4495	103,4832	20,8144	0,0000
Ошибка	116	576,7172	4,9717		
Общий	119	887,1667			

Таблица 7

Регрессионный вывод для семнадцатилетних мальчиков

Модель	R	R ²	R ² _a	Стандартная ошибка оценки
17 летних мальчиков	0,592	0,350	0,333	2,229

Затем необходимо рассмотреть значимость коэффициентов уменьшенной модели. В таблице 8 представлены вычисленные регрессионные коэффициенты для уменьшенной модели, с помощью которых получается регрессионная модель (3), используя соотношение (1). После этого выполняется проверка гипотезы для коэффициента регрессии. В таблице 8 T -статистики всех коэффициентов b_i больше чем T -критерия ($T_{табл} = 1,645$, $n = 120$, $\alpha = 0,05\%$). Следовательно, все коэффициенты являются статистически значимыми. Также все величины VIF меньше чем 5 и можно сделать вывод, что в уменьшенной модели отсутствует мультиколлинеарность.

Таблица 8

Коэффициенты регрессионной модели семнадцатилетних мальчиков

	коэффициенты	стандартные ошибки	T-статистики	P-value	Коллинеарность статистика	
					толерантность	VIF
Констант	18,0105	1,8457	9,7581	0,0000		
ЖЕЛ	0,0007	0,0003	2,5563	0,0119	0,966	1,035

ЧСС	-0,0298	0,0142	-2,1075	0,0372	0,998	1,002
Руфье	-0,6434	0,0976	-6,5956	0,0000	0,964	1,037

Для прогнозирования значения параметра $PWC_{170}/кг$ необходимо подставить в формулу (1) значения коэффициентов из таблицы 8:

$$\hat{y} = 18,0105 + 0,0007x_1 - 0,0298x_2 - 0,6434x_3, \quad (3)$$

3. Сравнение двух моделей

После этого, можно использовать метод (Partial F тест) для сравнения двух моделей. Вычисленная величина Partial F равно 0.5066. В F-таблице указывает F – критерию, которая равна 1.9588 ($\alpha = 0.05$, $df_1=9$, $df_2=107$) в уровни значимости $\alpha = 0.05(5\%)$. В этом случае необходимо рассмотреть следующие ситуации, если $H_0: F_{расч} < F_{табл}$ – уменьшенная модель лучше, $H_1: F_{расч} > F_{табл}$ – полная модель лучше. По результатам сравнения полной модели (2) с уменьшенной моделью (3) оказывается, что в анализе шестнадцатилетних мальчиков, уменьшенная модель лучше полной модели. Для сравнения двух моделей с помощью формулы (теста Multiple Partial F):

$$PartialF = \frac{(SSR_{полная} - SSR_{уменьшенная}) / q}{MSE_{полная}}$$

Где, SSR – сумма квадратов регрессии, MSE – средне-квадратическая остаточная, q – разница между количествами регрессоров из двух моделей.

4. Анализ остатков

Проанализированы остатки уменьшенной модели. Для проверки анализа остатков рассматривается стандартизированная остаточная гистограмма, в которой показаны остатки нормального распределения и общая форма является приемлемой. Среднее значение равно $-2,56 \cdot 10^{-16}$ и стандартное отклонение равно 0,987. На Рис. 6 приведены результаты проверки нормальности остатков.

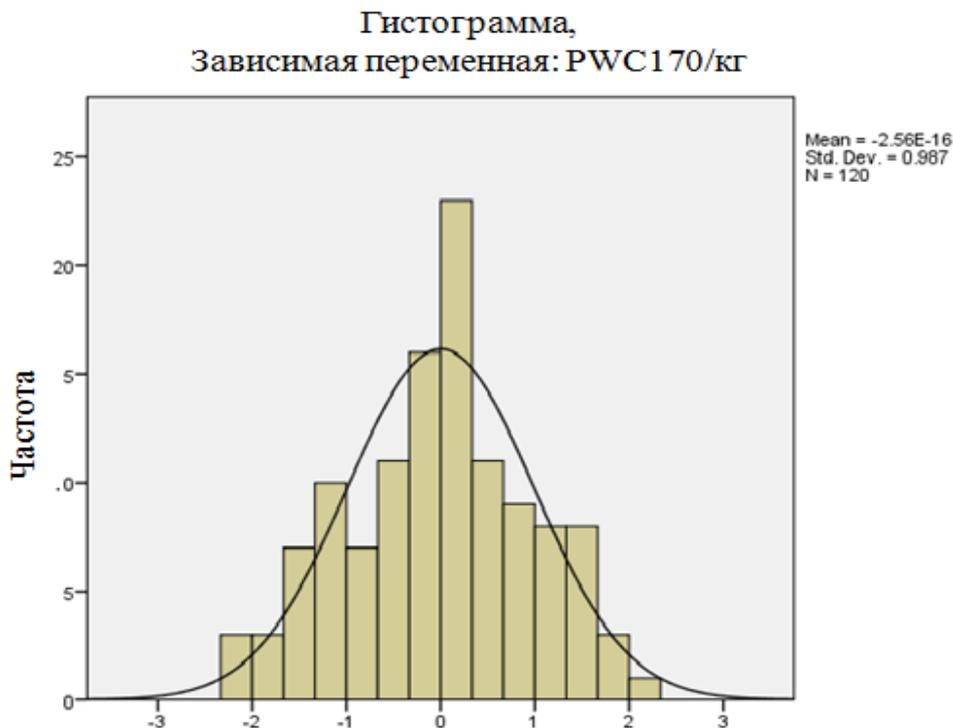


Рисунок 1. Гистограмма для стандартизированных остатков

После этого, рассмотрим проверку **гетероскедастичность** модели. Под гетероскедастичностью понимают неравные дисперсии остатков модели (проблема в дисперсии остатков). Рассмотрим точечную диаграмму (Рис. 2) для проверки гетероскедастичности остатков, в этой остаточной точечной диаграмме формы нарушения гетероскедастичности не видно, отсюда можно сделать вывод, что гетероскедастичность отсутствует в уменьшенной модели.

По математическому подходу, после вычисления, сравнивая величину Бройша Пагана с критерием хи-квадрат в таблице распределения хи-квадрата, получим величину Бройша Пагана = 4,21 и в таблице критическая величина хи-квадрата = 7,815. Очевидно, что величина Бройша Пагана меньше критерия хи-квадрат, что означает отсутствие гетероскедастичности в уменьшенной регрессионной модели.

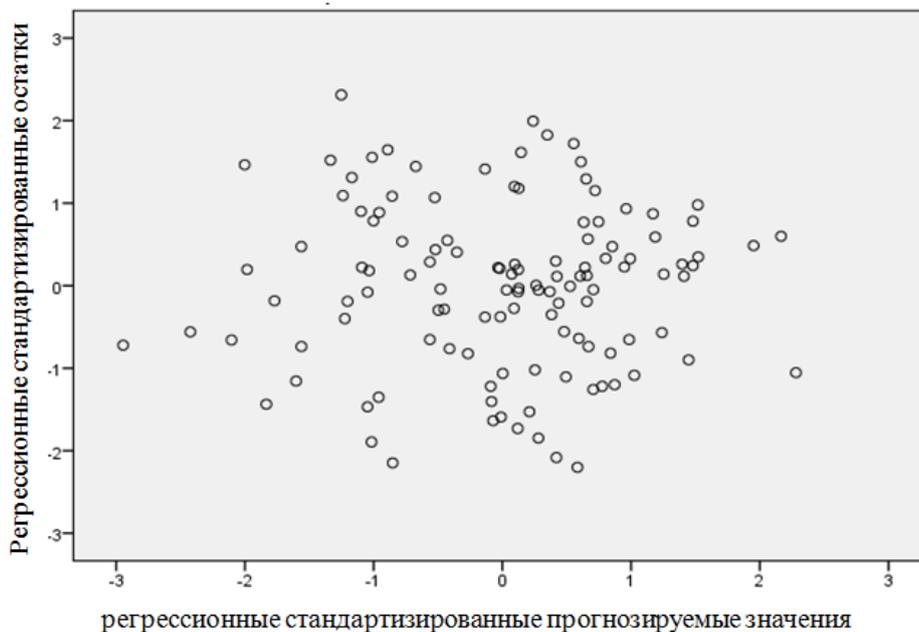


Рисунок 2. Точечная диаграмма остатков для 17 летних мальчиков

Затем, для обнаружения положительной автокорреляции можно использовать тест Дарбина Уотсона. **Автокорреляция** — статистическая взаимосвязь между последовательностями величин одного ряда, взятых со сдвигом, например, для случайного процесса со сдвигом по времени. По тесту Дарбина Уотсона $d = 1,863$ и в таблице указывается $d_{\text{нижняя}} = 1,584$, $d_{\text{верхняя}} = 1,665$. Поэтому величина Дарбина Уотсона $d > d_{\text{верхняя}}$ и можно считать, что положительная автокорреляция отсутствует в модели. Тест Дарбина Уотсона можно вычислить по следующей формуле:

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

где T -число наблюдений, e – остатки.

В результате получим, что после проверки гипотезы (F-тест, t-тест, и т.д.), модель множественной регрессии физической работоспособности семнадцатилетних мальчиков является полезной, и найдено применение для оценки состояния здоровья подростков. Другие регрессионные модели для восьми до семнадцатилетних юношей и девушек также позволяют получить аналогичные результаты.

5. Разработка программного обеспечения

После получения результатов модели пятнадцатилетних девушек разработана программа на языке C# с использованием Microsoft Visual Studio 2012. При использовании данной программы потребитель должен вводить вводные данные с помощью клавиатуры. Соответственно программа рассчитает физическую работоспособность. После расчета программа дает предсказанную физическую работоспособность на экране. Расчет функции для пятнадцатилетних девушек по программе выглядит следующим образом:

```
private void female15()
{
    //define coefficient of the model of 15 year old girls
    double[] female15 = new double[7];
    female15[0] = 18.6021;
    female15[1] = -0.0347;
    female15[2] = -0.0096;
    female15[3] = -0.0516;
    female15[4] = 0.0584;
    female15[5] = 0.0614;
    female15[6] = -0.4216;
    // changing string to double from input data
    string strHR = txtHR.Text;
    string strWeightHeight = txtWeightHeight.Text;
    string strCatchBall = txtCatchBall.Text;
    string strRulerTest = txtRulerTest.Text;
    string strBelly = txtBellyPress.Text;
    string strSitUps = txtSitUps.Text;

    double HR = double.Parse(strHR, System.Globalization.CultureInfo.InvariantCulture);
    double weightHeight = double.Parse(strWeightHeight, System.Globalization.CultureInfo.InvariantCulture);
    double catchBall = double.Parse(strCatchBall, System.Globalization.CultureInfo.InvariantCulture);
    double rulerTest = double.Parse(strRulerTest, System.Globalization.CultureInfo.InvariantCulture);
    double Belly = double.Parse(strBelly, System.Globalization.CultureInfo.InvariantCulture);
    double sitUps = double.Parse(strSitUps, System.Globalization.CultureInfo.InvariantCulture);
    // assign model of 15 yrs old girls
    double pwc = female15[0] + (female15[1] * HR) + (female15[2] * weightHeight) +
        (female15[3] * catchBall) + (female15[4] * rulerTest) + (female15[5] * Belly) + (female15[6] * sitUps);
    // showing Output PWC
    label11.Text = label11.Text + pwc;
}
```

Рисунок 3. Расчета функции для пятнадцатилетних девушек

В программе пользователь может выбрать модель для возраста и пола. После этого пользователю необходимо ввести данные в соответствующее поле, и потом нажать кнопку "Calculate" и программа покажет вычисленный результат прогнозируемой физической работоспособности. Программа отображает следующим образом:

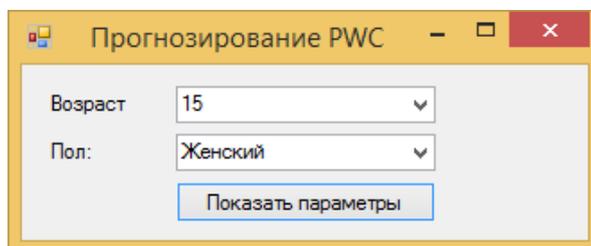


Рисунок 4. выбрать возраст и пол в программе

Parameter	Value
ЖЕЛ	
ЧСС	98
АД-С	
АД-Д	
Гипокс.	
Кетле	329
Гибк.	
Коорд	5
ЗРД	18
Отжим	
Пресс	22
Руфье	12

Calculate

PWC = 9,1279

Рисунок 5. вводить данные для расчета

Заключение

1. Для прогнозирования физической работоспособности человека выбран метод множественного регрессионного анализа статистики, который позволяет проводить анализ многофакторных статистических моделей.
2. Разработаны регрессионные модели для прогнозирования физической работоспособности человека в возрасте от восьми до семнадцати лет.
3. Разработано программное обеспечение для прогнозирования физической работоспособности человека в возрасте от восьми до семнадцати лет.
4. Определены значимые параметры, с помощью которых быстро и эффективно можно оценить физическую работоспособность человека в возрасте от восьми до семнадцати лет.
5. По результатам регрессионного анализа, регрессионные модели человека в возрасте от восьми до семнадцати лет являются статистически значимыми и могут быть использованы при оценке состояния здоровья.
6. Найдены значимые параметры для оценки физической работоспособности человека в возрасте от восьми 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; Вып. 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), - 671 p.
3. Statistical Models: Theory and Practice, David A. Freedman, Cambridge University Press (2005), - 414 p.
4. Modeling and interpreting interactive hypotheses in regression analysis, Cindy D. Kam and Robert J. Franzitsi Jr., University of Michigan Press (2009), - 168 p.
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, - 366 p.
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), - 259 p.
8. Народный СпортПарк. Режим доступа: <http://sportpark.ru/> (дата обращения 15.12.2015).
9. President of Russia, Address to Federal Assembly Russia. Режим доступа: http://archive.kremlin.ru/eng/speeches/2005/04/25/2031_type70029type82912_87086.shtml (дата обращения 15.12.2015).

Kyi Thar Soe

Bauman Moscow State Technical University «BMSTU», Russia, Moscow
E-mail: kyithar82@gmail.com

Software development for the prediction of human physical working capacity based on regression analysis

Abstract. The paper presents results of research for the mathematical model of the physical working capacity 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 constructed for each age group. After performing multiple regression analysis obtained multiple regression models that can predict the physical working capacity for human aged eight to seventeen years. Article shows the results of the analysis and development of the regression models for the boy of seventeen year old. Also presents the hypotheses testing, ie checking the significance of the model, the significance of the coefficients, heteroscedasticity, autocorrelation, multicollinearity and normality of residuals. In conclusion, developed the regression models and software for the prediction of physical working capacity, method selected multiple regression analysis in statistics, which allows the analysis of multivariate statistical models. Defined the relevant parameters for the regression equations to assess the physical health of girls between the ages of fourteen to seventeen years quickly and efficiently.

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

REFERENCES

1. Matematicheskaya statistika: Ucheb. Dlya vuzov / V.B. Goryaninov, I.V. Pavlov, G.M. Krishchenko. -2-e izd., stereotip. – M.: Izd-vo MGTU im. N.E. Baumana, 2002. – 424 s. (Ser. Matematika v tekhnicheskoy 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), - 671 p.
3. Statistical Models: Theory and Practice, David A. Freedman, Cambridge University Press (2005), - 414 p.
4. Modeling and interpreting interactive hypotheses in regression analysis, Cindy D. Kam and Robert J. Franzitsi Jr., University of Michigan Press (2009), - 168 p.
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, - 366 p.
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), - 259 p.
8. Narodnyy SportPark. Rezhim dostupa: <http://sportpark.ru/> (data obrashcheniya 15.12.2015).
9. President of Russia, Address to Federal Assembly Russia. Rezhim dostupa: http://archive.kremlin.ru/eng/speeches/2005/04/25/2031_type70029type82912_87086.shtml (data obrashcheniya 15.12.2015).