

КОНСТРУИРОВАНИЕ МОДЕЛЕЙ И ПОСТРОЕНИЕ КРАТКОСРОЧНОГО ПЕРСПЕКТИВНОГО ПРОГНОЗА ВЫПУСКА СПЕЦИАЛИСТОВ

А.А. Гаврилова, П.А. Голованов, Н.В. Дилигенский, Е.П. Тупоносова

Самарский государственный технический университет
443100, г. Самара, ул. Молодогвардейская, 244
E-mail: a.a.gavrilova@mail.ru

Проанализировано влияние приема абитуриентов и демографических факторов на выпуск специалистов СамГТУ по годам и специальностям. Построены математические модели, позволяющие краткосрочное прогнозирование выпуска специалистов СамГТУ.

Ключевые слова: *системный подход, прием и выпуск специалистов, математическое моделирование, модель, регрессионный анализ, прогнозирование.*

Проводимая государством стратегия экономии в последние годы приводит к ежегодному сокращению бюджетного приема в вузы примерно на 3,8%. При этом проводится реформа в системе среднего образования и добавляются демографические проблемы, приводящие к сокращению выпуска школьников и соответственно контингента поступающих в вузы. Поэтому актуальными становятся анализ динамики приема и прогнозирование выпуска специалистов, позволяющие определить наиболее вероятные варианты развития событий и возможности воздействия на ситуацию.

В качестве объекта исследования рассматривается Самарский государственный технический университет (СамГТУ). В качестве входного фактора примем поступление абитуриентов, в качестве выходного – выпуск специалистов очного обучения в период с 1998 по 2010 гг. Поставим задачу построения модели подготовки инженеров на основании статистических данных приема и выпуска СамГТУ.

Основными целями использования математического моделирования являются: выявление функциональных соотношений – отыскание количественных зависимостей между входным фактором модели (прием абитуриентов) и выходным (выпуск специалистов), прогнозирование – оценка будущего поведения объекта при некоторых предполагаемых сочетаниях влияющих воздействий и отыскание способов управления объектом.

Основными стадиями построения и применения математических моделей социально-производственных систем являются следующие: содержательная постановка задачи, математическая формализация задачи, решение сформулированной задачи, анализ и интерпретация полученных результатов и верификация модели [1].

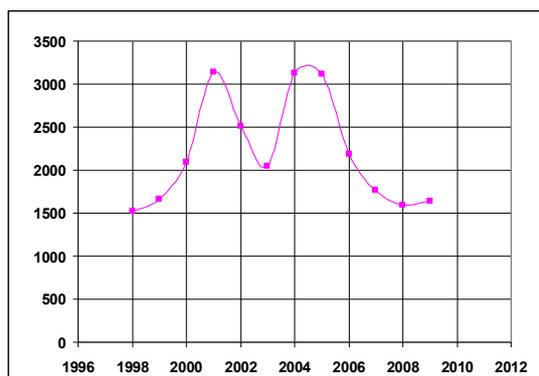
Проведем анализ статистических данных приема абитуриентов дневного отделения и выпуска специалистов по годам и специальностям СамГТУ с 1998 по 2009 гг. (рис. 1, рис. 2).

Проанализируем численность приема абитуриентов в этот период (см. рис. 1). Процесс приема студентов является неравномерным, характеризуется периодами

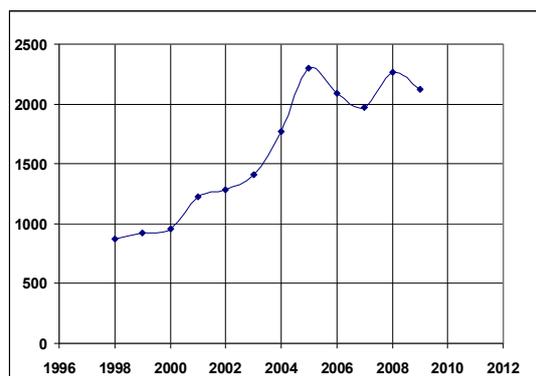
*Анна Александровна Гаврилова – к.т.н., доцент.
Павел Александрович Голованов – старший преподаватель.
Николай Владимирович Дилигенский – д.т.н., профессор.
Елена Павловна Тупоносова – аспирант.*

подъема в 2001, 2004 и 2005 гг. и спада в 2003, 2006-2009 гг. Контингент вуза почти удвоился с 1998 по 2001 г., прием вырос на 97%. Далее отмечался непродолжительный спад в течение 2 лет – до 2003 года, который составил 49%. С 2005 до 2009 г. продолжительное снижение приема привело к уровню 1998 г. При максимальном приеме в вуз поступило порядка 3100 абитуриентов; минимальный прием был в 1998 г. – около 1500 человек.

В целом исходные статистические данные приема абитуриентов соответствуют численным показателям выпуска школ Самарской области [2]. От общего выпуска школ в СамГТУ поступают около 8% абитуриентов. Небольшой прирост может происходить за счет выпускников школ прошлых лет, учреждений профессионального образования и средних специальных учебных заведений.



Р и с . 1. Прием студентов



Р и с . 2. Выпуск специалистов

Анализ данных выпуска специалистов, представленный на рис. 2, показывает, что их численность возрастала до 2005 г., рост на 7% в год в 1998-2002 гг. сменился значительным ростом – до 60% в 2003-2005 гг. Максимальное количество специалистов – 2300 человек – вуз выпустил в 2005 и 2008 гг., что соответствует максимальному приему в 2001 и 2004 гг.

Таким образом, процессы приема и выпуска являются коррелированными. В соответствии с пятилетним сроком обучения специалистов максимальный выпуск 2005-2006 гг. соответствует наибольшему приему абитуриентов в 2001 г. Выпуск 2007 года (1970 человек) соответствует минимальной точке приема в 2003 г.

Средний отсев студентов за 5 лет обучения в годы максимального приема составил 26,7% и был связан с отчислением по неуспеваемости, переводом на заочный факультет и перераспределением по факультетам. В потоках, соответствующих малому приему, отсев составил 14% от числа принятых.

Для выявления статистической взаимосвязи процессов приема и выпуска специалистов в вузе по исходным данным построим математическую модель с возможностью краткосрочного прогноза выпуска специалистов СамГТУ на период до 2013 г.

Качество модели определяется ее близостью к исследуемому процессу и точностью. Адекватность характеризуется наличием и учетом определенных статистических свойств, а точность – степенью соответствия фактическим данным.

Оценка адекватности исходных статистических данных и модели производится по следующим характеристикам: коэффициент корреляции r_{xy} ; коэффициент детерминации R^2 ; среднее квадратическое отклонение σ ; критерий Фишера; стандартная

ошибка разности средних арифметических σ_{x-y} ; критерий Стьюдента; критерий Дарбина – Уотсона d ; коэффициент автокорреляции r_a [3].

Коэффициент корреляции, характеризующий связь между входными (x) и выходными (y) случайными величинами, вычисляется по формуле

$$r_{x,y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \cdot \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, x_i – количество абитуриентов, принятых в вуз; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, y_i – выпуск специалистов.

Коэффициент корреляции характеризует тесноту, или силу, связи между параметрами: приемом студентов и выпуском специалистов. Абсолютная величина коэффициента корреляции не превышает единицы $|r_{xy}^*| \leq 1$: чем ближе r_{xy} к единице, тем ближе стохастическая зависимость между переменными.

Для измерения мультиколлинеарности – попарной корреляционной зависимости между факторами – используют коэффициент множественной детерминации:

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

Коэффициент R^2 показывает, какая доля дисперсии результативного признака объясняется влиянием объясняющих переменных. Если связь отсутствует, то R^2 равно 0. Коэффициент детерминации не определяет, увеличивается ли или уменьшается выход с ростом входа. Функциональная связь возникает при значении, равном 1, а отсутствие связи – 0.

Среднее квадратическое отклонение σ является показателем надежности средней: чем меньше σ , тем лучше средняя арифметическая отражает всю статистическую совокупность. Эта величина характеризует разброс данных. Для оценки величины σ в выборке используют формулу

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad (3)$$

где n – число наблюдений.

Статистическая значимость коэффициента детерминации R^2 проверяется нулевой гипотезой критерия Фишера:

$$F_{\text{эмн}} = \frac{\sigma_x^2}{\sigma_y^2}, \quad (4)$$

где σ_x, σ_y – дисперсии первой (приема) и второй (выпуска) выборки соответственно.

По таблице находим критические значения критерия Фишера по величинам k_1 и k_2 . Если $F_{эмп} > F_{крит}$, то можно утверждать, что нулевая гипотеза принимается, уравнение статистически значимо и по степени однородности показателя входных и выходных величин различий между выборками нет [3].

Для нахождения вероятности того, что оба средних значения в выборке относятся к одной и той же совокупности, применяют критерий Стьюдента (t-критерий):

$$t_{эмп} = \frac{\bar{x} - \bar{y}}{\sigma_{x-y}}, \quad (5)$$

где стандартная ошибка разности средних арифметических определяется по формуле

$$\sigma_{x-y} = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{(n-1)n}}. \quad (6)$$

Подсчет числа степеней свободы осуществляется по формуле $k = n_1 + n_2 - 2$. Если при сравнении полученного значения $t_{эмп}$ с теоретическим t-распределением Стьюдента $t_{эмп} < t_{крит}$, то гипотеза принимается, в противном случае нулевая гипотеза отвергается и принимается альтернативная гипотеза.

Критерий Дарбина – Уотсона используется для оценки статистической зависимости переменных:

$$d = \frac{\sum_{i=2}^n ((y_i - y_{iT}) - (y_{i-1} - y_{iT-1}))^2}{\sum_{i=1}^n (y_i - y_{iT})^2}, \quad (7)$$

где y_i, y_{i-1} – действительный выпуск; y_{iT}, y_{iT-1} – расчетный выпуск; n – объем выборки. Допустимый интервал для критерия Дарбина – Уотсона изменяется в пределах $0 < d < 4$.

Коэффициент автокорреляции рассчитывается по формуле

$$r_a = 1 - \frac{d}{2}, \quad (8)$$

где d – статистика Дарбина – Уотсона.

Данный коэффициент показывает наличие автокорреляции – взаимосвязаны ли последовательные элементы рядов данных. Если $r_a \leq r_{aT}$, то уровни динамического ряда независимы [4].

На первом этапе в качестве аппарата исследования выбираем наиболее простые регрессионные модели. Построим линейную однофакторную модель методом регрессионного анализа в следующем виде:

$$y = A_0 + A_1x, \quad (9)$$

где y – выпуск специалистов; x_1 – количество студентов, принятых в вуз; A_1 – коэффициент регрессии; A_0 – свободный член уравнения.

Параметры модели (9) идентифицируются методом наименьших квадратов на

основе минимизации отклонения от точек исходного временного ряда. К исходному ряду с целью устранения влияния случайных данных применим метод сглаживания ряда по трем точкам. В результате получим уравнение регрессии в следующем виде:

$$y = 485,7 + 0,58x. \quad (10)$$

В табл. 1 приведены значения показателей качества модели (10), на рис. 3 показано сравнение расчетных и исходных данных.

Таблица 1

r_{xy}	R^2	σ_x	F	σ_{x-y}	t	d	r_a
0,93	0,85	0,668	3,063	0,27	1,83	1,41	0,29

Аппроксимативные свойства модели (10), определяемые коэффициентом детерминации R^2 , характеризуют взаимную корреляцию фактических данных и результатов расчетов, но остальные рассчитанные коэффициенты характеризуют недостаточные описательные свойства модели, а коэффициент автокорреляции неудовлетворителен и составляет $r_a = 0,298$. Для периода 2003-2005 гг. модельные расчеты практически точно совпадают с исходными данными. Однако в целом имеют место значительные расхождения модели и исходных данных. Указанное описание не является адекватным анализируемому процессу, прогноз по нему производиться не будет.

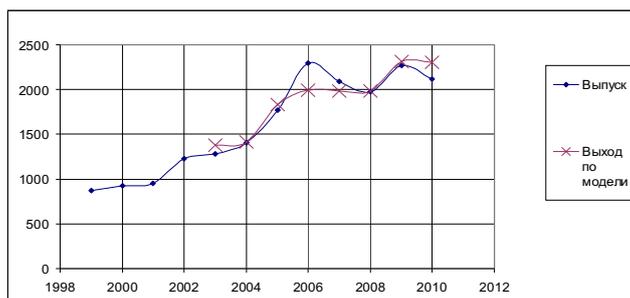


Рис. 3. Линейная сглаженная однофакторная модель выпуска специалистов

Для улучшения описательных свойств модели увеличим количество факторов, определяющих выпуск специалистов, учтем влияние наиболее значимых факультетов. Контингент специалистов на факультете автоматике и информационных технологий (ФАИТ) составляет 15,9%, на нефтетехнологическом (НТФ) – 14,4% и инженерно-экономическом (ИЭФ) – 10,9%, что в наибольшей степени определяет подготовку специалистов вуза в целом. Построим линейную двухфакторную модель исследования влияния трех факультетов на общий выпуск СамГТУ:

$$y = A_0 + A_1x_1 + A_2x_2, \quad (11)$$

где x_1 – общий прием абитуриентов в вуз, исключая прием студентов на три факультета (ФАИТ, НТФ и ИЭФ); x_2 – суммарный прием студентов на три факультета.

Параметры модели (11) идентифицируем методом наименьших квадратов[3]:

$$y = 775 + 0,3x_1 + 0,67x_2. \quad (12)$$

Статистический анализ аппроксимативных свойств модели (12) представлен в табл. 2, на рис. 4 показано сравнение данных по модели и исходных.

Таблица 2

r_{xy}	R^2	σ_x	F	σ_{x-y}	t	d	r_a
0,82	0,82	0,7	3,4	0,22	2,23	3	-0,53

Полученные коэффициенты близки по своим значениям к модели (10), но разница между фактическим выпуском и модельным расчетом составляет более 18%. Стохастическая зависимость между переменными стала хуже, стандартная ошибка разности средних арифметических уменьшилась на 18%. По критерию Дарбина – Уотсона автокорреляция остатков изменилась с положительного значения на отрицательное. Это означает, что выбранная функция не отражает реальную взаимосвязь исследуемых величин. Прогноз по данной модели производить не будем.

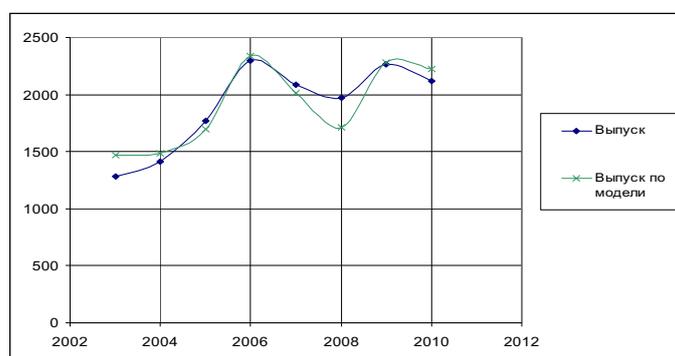


Рис. 4. Линейная двухфакторная модель выпуска специалистов

В качестве аппроксимирующей функции рассмотрим квадратичный полином. Построим модель в следующем виде:

$$y = A_0 + A_1x + A_2x^2, \quad (13)$$

где A_1, A_2 – коэффициент регрессии; A_0 – свободный член уравнения; x – количество студентов, принятых в вуз; y – выпуск специалистов.

По методу наименьших квадратов получаем следующее описание:

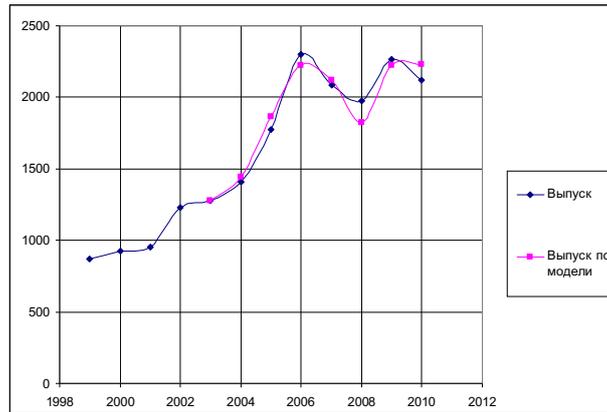
$$y = -0,0005x^2 + 2,58x - 1665. \quad (14)$$

Статистический анализ аппроксимативных свойств модели (14) приведен в табл. 3, на рис. 5 показано сравнение рассчитанных и исходных данных.

Аппроксимативные свойства модели (14) характеризуют высокие описательные свойства модели, но значение критерия Дарбина – Уотсона $d = 2,7$ велико и находится близко к верхней границе. Полученные коэффициенты свидетельствуют о том, что не учтены основные влияющие факторы и модель не обладает удовлетворительными прогнозными свойствами.

Таблица 3

r_{xy}	R^2	σ_x	F	σ_{x-y}	t	d	r_a
0,93	0,86	0,67	3,06	0,27	1,82	2,7	-0,37



Р и с . 5. Квадратичная модель выпуска специалистов

Для устранения случайных отклонений произведем сглаживание исходных данных. При этом идентификация коэффициентов модели на основе сглаженных статистических данных дала следующие результаты [4]:

$$y = -0,0002x^2 + 1,64x - 742 . \quad (15)$$

В табл. 4 приведены показатели качества модели (15), на рис. 6 показано сравнение рассчитанных и исходных данных.

Таблица 4

r_{xy}	R^2	σ_x	F	σ_{x-y}	t	d	r_a
0,94	0,86	0,46	1,94	0,2	2,31	1,8	0,1

При сглаживании исходных данных по методу скользящего среднего по 3 точкам параметры модели существенно улучшились. Полученные коэффициенты характеризуют высокие описательные свойства модели, позволяющие строить прогнозы.

По прогнозу на 3 года по модели (15) следует, что выпуск специалистов будет сокращаться в среднем на 26% по сравнению с 2009 г. В 2013 г. выпуск составит порядка 1300 человек, что соответствует выпуску 2003 г.



Р и с . 6. Сглаженная квадратичная модель прогнозирования выпуска специалистов до 2013 г.

Представленные выше результаты свидетельствуют о том, что не учтены некоторые влияющие факторы и модель не обладает хорошими прогнозными свойствами. Добавим фактор количества выпускников школ [2]. Применим модель в виде производственной функции (ПФ) Кобба – Дугласа [1] по сглаженным исходным данным и на ее основе построим математическую модель:

$$Y = AK^{\alpha}V^{\beta}, \quad (16)$$

где Y – общая численность выпускников вуза в расчетный год; K – количество выпускников школы за 5 лет до расчетного года; V – количество абитуриентов вуза; α, β – функции эластичности; A – масштабный коэффициент.

В результате обработки исходного ряда получаем: $A = 0,002, \alpha = 0,12, \beta = 1,28$.

$$Y = 0,002K^{0,12}V^{1,28}. \quad (17)$$

Статистический анализ аппроксимативных свойств модели (17) дал следующие результаты (табл. 5).

Таблица 5

r_{xy}	R^2	σ_x	F	σ_{x-y}	t	d	r_a
0,95	0,87	0,68	2,9	0,21	2,23	2,43	-0,21

Модель обладает высокими прогнозными свойствами – критерий Дарбина – Уотсона равен $d = 2,43$. Между модельными расчетами и реальными данными практически отсутствует разница. Построенная по расчетным данным модель в соответствии с критерием Дарвина – Уотсона позволяет прогнозировать выпуск специалистов [5].

На рис. 7 показана модель прогнозирования с помощью ПФ выпуска специалистов до 2013 г.

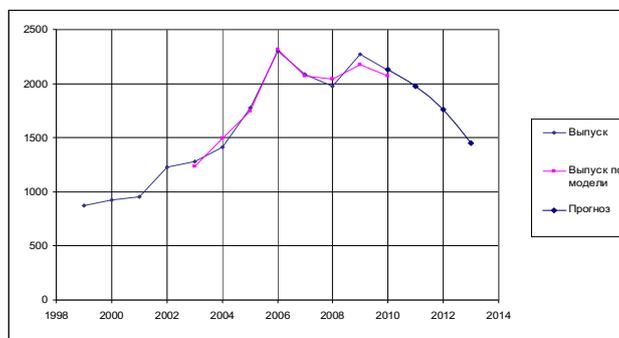


Рис. 7. Модель прогнозирования с помощью ПФ выпуска специалистов до 2013 г.

Модель (17) показывает, что по прогнозу на 3 года выпуск специалистов будет сокращаться в среднем на 17% по сравнению с 2009 г. Это связано с понижением рождаемости в период с 1993 по 2000 г. в среднем на 3 тысячи детей в год.

В табл. 6 приведены показатели качества всех построенных моделей.

Проведенные исследования показывают, что не все построенные модели позволяют прогнозировать выпуск специалистов. Лучшими прогнозными свойствами обладают квадратичная модель со сглаженными данными и математическая модель на основе производственной функции Кобба – Дугласа. При этом введение в модель дополнительного фактора – выпуска школ – позволило более адекватно описать процесс и повысить показатели качества модели.

Таблица 6

Модель	r_{xy}	R^2	σ_x	F	σ_{x-y}	$t_{эмл}$	d	r_a
10	0,93	0,85	0,67	3,06	0,27	1,83	1,41	0,29
12	0,82	0,82	0,7	3,4	0,22	2,23	3	-0,53
13	0,92	0,9	0,62	3,2	0,2	1,93	1,5	0,3
15	0,93	0,86	0,67	3,06	0,27	1,82	2,7	-0,37
16	0,94	0,86	0,46	1,94	0,2	2,31	1,8	0,1
18	0,95	0,87	0,68	2,9	0,21	2,23	2,4	-0,21

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Дилигенский Н.В., Гаврилова А.А., Цапенко М.В. Построение и идентификация математических моделей производственных систем: Учеб. пособие. – Самара: Офорт, 2005. – 126 с.
2. Самарская область. Контингент, прием, выпуск студентов и учащихся с 1996 по 2004 г. <http://labourmarket.ru/Pages/Levels/volga/samara-region/index.htm>
3. Бережная Е.В., Бережной В.И. Математические методы моделирования экономических систем: Учеб. пособие. – М.: Финансы и статистика, 2002. – 368 с.
4. Дилигенский Н.В., Цапенко М.В., Гаврилова А.А. Математические модели управления производственно-экономическими системами: Учеб. пособие / Самар. гос. техн. ун-т. – Самара, 2005. – 112 с.
5. Гаврилова А.А., Голованов П.А., Дилигенский Н.В., Тупоносова Е.П. Системный анализ и прогнозирование подготовки специалистов с высшим техническим образованием // Известия Самарского научного центра РАН.

Статья поступила в редакцию 24 мая 2011 г.

UDC 519.5 620.9

DESIGN OF MODELS AND MAKING SHORT-TERM PROGNOSIS OF EXPERTS DRADUATION

A.A. GavriloVA, P.A. Golovanov, N.V. Diligenskii, E.P. Tuponosova

Samara State Technical University
244, Molodogvardeyskaya st., Samara, 443100

The influence of enrolle admission and demographic factors on SSTU graduation for years and professions is analyzed. Mathematical models for graduation forecasting are build.

Keywords: *system approach, the reception and production professionals, mathematical modeling, the regression analysis, the model prediction.*

A.A. GavriloVA – Candidate of Technical Sciences, Associate Professor.

P.A. Golovanov – Senior Lecturer.

N.V. Diligenskii – Doctor of Technical Sciences, Professor.

E.P. Tuponosova – Postgraduate student.