

Прогнозирование себестоимости документов в управлении информационными системами

© 2012 Г.Н. Исаев

кандидат технических наук, профессор

Институт сервиса, г. Москва

Филиал Российского государственного университета туризма и сервиса

E-mail: georg.isaev@mail.ru

Автором в контексте проектного финансирования рассмотрена разработка модели прогнозирования себестоимости получения выходных документов информационных систем. Дано описание компьютерной обработки данных для экспериментальной проверки адекватности модели.

Ключевые слова: себестоимость документов, прогнозирование, моделирование, проектное финансирование, управление качеством, информационные системы.

Инвестирование проектов и эксплуатации информационных систем (ИС) различного класса и назначения в настоящее время составляет значительную долю в общей структуре затрат на всех уровнях управления экономикой страны. Проектное финансирование сравнительно мало-прибыльных ИС осуществляется в основном по форме “финансирование с полным регрессом на заемщика”¹. Основные риски падают на заемщика. Поэтому предъявляются повышенные требования к срокам окупаемости проекта, возврата средств, к качеству ИС.

Одним из эффективных направлений в минимизации сроков является управление качеством ИС на всех стадиях жизненного цикла - исследования, проектирования, построения и функционирования². В управлении качеством ИС решается комплекс задач по измерению и оценке качества ИС не только со стороны функциональной, но и со стороны экономической составляющей на основе системы показателей, например, времени, трудоемкости, стоимости, себестоимости получения выходных документов как формы финишной продукции ИС и др. С позиций эконометрии и квалиметрии показатели рассматриваются как иерархическая система свойств, отражающая качество ИС, т.е. единичные, групповые, интегральные, комплексные, обобщенные³. Оценку сроков окупаемости проекта и возврата средств можно выполнить с учетом прогноза значений показателей, в частности, себестоимости выходных документов ИС. На практике себестоимость документов значительно снижается от наличия дефектов, возникающих на различных участках и этапах ИС, например, искажения значений показателей, отсутствия (пропусков) значений показателей входных документов, нарушения регламентных сроков представления документов и др. Для об-

наружения и исправления дефектов расходуются значительные ресурсы. С учетом этого условия при расчете себестоимости целесообразно использовать функциональную зависимость между значимостью (“весомостью”) дефектов технологии ИС и значением себестоимости выходных документов ИС. Указанную зависимость можно определить на основе инструментов математической статистики, например, регрессионного анализа⁴. Выбор вида функции зависимо-го показателя себестоимости от независимых показателей должен быть выполнен так, чтобы получаемая при этом линейная зависимость была бы лучшей аппроксимацией функциональной зависимости.

Реализация модели прогнозирования себестоимости выходных документов ИС может быть выполнена посредством множественной линейной регрессии

$$Y_i = a_0 + A'x_i + \varepsilon_i = a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_q x_{iq} + \varepsilon_i, \quad (1)$$

где Y_i - зависимая (прогнозируемая) переменная - себестоимость документа;

x_i - независимые (прогнозирующие) переменные (значения стоимости обнаружения и исправления дефектов по достоверности, полноте, своевременности);

a_0 - свободный член регрессии;

A' - вектор оценок коэффициентов линейной регрессии;

ε_i - случайные величины (совокупность латентных случайных факторов).

Оценка параметров a_0 , A' производится методом наименьших квадратов, т.е. из условия ми-

нимума суммы квадратов отклонений значений Y_i :

$$\Delta^2 = \sum_{i=1}^n (y_i - A'x_i - a_0)^2. \quad (2)$$

Это приводит к системе нормальных уравнений

$$\begin{cases} \hat{A} = S^{-1}\hat{C}_{yx} \\ a_0 = m_y - \hat{A}' M_x, \end{cases} \quad (3)$$

где $\hat{A} = (\hat{a}_1, \dots, \hat{a}_q)'$ - вектор оценок коэффициентов линейной регрессии;

\hat{a}_0 - свободный член уравнения регрессии;

S^{-1} - обратная матрица ковариаций между переменными x_1, \dots, x_q ;

\hat{C}_{yx} - вектор оценок ковариаций между переменными y и переменными x_1, \dots, x_q ;

m_y - оценка среднего значения y ;

M_x - вектор средних значений переменных x_1, \dots, x_q .

Для оценки переменных регрессии необходимо получить исходные данные. В нашем случае они могут быть представлены в виде матрицы фиксированных данных. Матрица фиксированных данных имеет размерность $n \times m (m=q+1, x_m=Y)$ и представляет собой выборку m -мерных объектов $X=(x_{1, \dots, q, \dots, m})$. По условиям задачи необходимо иметь матрицу по себестоимости. Расчет значений целесообразно выполнить исходя из зависимости снижения себестоимости от уменьшения стоимости обнаружения и исправления дефектов на 1 %. Исходя из условий проверки адекватности модели, а также практических соображений использования результатов работы, целесообразно представить матрицу исходных данных, состоящую из 20 строк каждая, т.е. просчитать зависимость до 20 %.

Матрица фиксированных данных для расчета регрессионной зависимости дефектов обработки и себестоимости документов определяется в следующем порядке. Оцениваем математическое ожидание дефекта обработки по формуле

$$x_j^c = \overline{x_j^c} \cdot P_j, \quad (4)$$

где x_j^c - оценка математического ожидания стоимости обнаружения и исправления одного дефекта по j -й переменной матрицы;

$\overline{x_j^c}$ - среднее выборочное значение стоимости обнаружения и исправления одного дефекта по j -й переменной, полученной ранее в результате обработки дефектов;

P_j - относительная частота j -й переменной, приходящаяся на один документ.

Затем оцениваем общую стоимость дефектов C_j^d по формуле

$$C_j^d = x_j^c \cdot V, \quad (5)$$

где V - объем обрабатываемой документации (измеряемый в количестве документов, показателей, символов).

Тогда совокупная стоимость обнаружения и исправления дефектов C^d может быть определена по формуле

$$C^d = \sum_{j=1}^q C_j^d. \quad (6)$$

Определим общую стоимость дефектов C_{jp}^d по j -й переменной при условии снижения стоимости на p процентов по формуле

$$C_{jp}^d = C_j^d - [(C_j^d / 100) \cdot p]. \quad (7)$$

Значение C_{jp}^d записываем в соответствующие графы матрицы фиксированных данных. Для определения переменной - себестоимости обработки одного документа - необходимо получить совокупную стоимость обнаружения и исправления дефектов при условии снижения стоимости дефектов на p процентов по формуле

$$C_p^d = \sum_{j=1}^q C_{jp}^d, \quad (8)$$

где C_p^d - совокупная стоимость обнаружения и исправления дефектов при условии снижения стоимости на p процентов.

Определим нормированную стоимость обработки документации, т.е. стоимость при условии отсутствия дефектов по формуле

$$C^0 = C^N \cdot V, \quad (9)$$

где C^0 - общая нормированная стоимость обработки документации ИС;

C^N - нормированная стоимость обработки одного документа.

Теперь определим общую фактическую стоимость обработки документации

$$C_p^f = C_p^d + C^0, \quad (10)$$

где C_p^f - общая фактическая стоимость обработки при условии снижения стоимости дефектов на p процентов.

Тогда значение зависимой переменной определяем по формуле

$$C_p^y = C_p^f / V, \quad (11)$$

где C_p^y - значение себестоимости обработки одного документа при условии снижения стоимости дефектов обработки на p процентов.

Записав в соответствующую позицию значение

C_p^y матрицы, проводим вычисления переменных следующей строки.

Работоспособность и адекватность модели была проверена посредством планирования вычислительного эксперимента с разработкой соответствующих технологических документов и аппаратно-программного комплекса⁵. Сведения о дефектах регистрировались в "Ведомости выявленных дефектов". Данные ведомости дефектов были введены в компьютер и обработаны программой кластер-анализа статистической структуры дефектов посредством пакета прикладного статистического анализа⁶. В результате кластеризации сформированы три класса дефектов - по достоверности, полноте и своевременности. В соответствии с моделью (формулы 1-11) на основе статистических оценок получена матрица фиксированных данных по себестоимости

документов ИС, представленная фрагментом (см. табл. 1). В гр. 1 указаны значения процентов снижения значений переменных, которые указаны в гр. 2-4, значения прогнозируемой переменной - в гр. 5. В табл. 1 предсказывающие переменные измеряются в рублях, прогнозируемая переменная - в рублях за 1 документ. Обработка матрицы выполнена посредством программы регрессионного анализа. В результате обработки данных матрицы были получены коэффициенты регрессии и оценочные величины по модели себестоимости (см. табл. 2), а также график зависимости себестоимости от дефектов ИС (рис. 1).

Коэффициент множественной корреляции - 0,99917.

Коэффициент детерминации - 0,99833.

Приведенная (несмещенная) оценка коэффициента детерминации - 0,99804.

Стандартная ошибка вычислений - 0,020831.

F -значение статистики Фишера для проверки нулевой гипотезы - 3392.

Уровень значимости (P -значение) нулевой гипотезы - 0,0000.

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>.

По данным табл. 2 составлено уравнение множественной линейной регрессии по себестоимости

$$y^c = 3,663 + 0,3924 x_1 + 0,0002067 x_2 - 0,02332 x_3.$$

В правой части уравнений расположены слева направо соответственно свободный член регрессии

Таблица 1. Матрица фиксированных данных по себестоимости

Значение %	Достоверность, руб.	Полнота, руб.	Своевременность, руб.	Себестоимость, руб./док.
1	2	3	4	5
0	15776,72	109383,76	266067,02	11,24
1	15618,95	108289,93	263406,06	11,17
2	15461,18	107196,10	260745,68	11,00
3	15303,42	106102,27	258085,01	11,02
4	15145,65	105008,44	255424,34	10,94
5	14987,88	103914,61	252763,67	10,86
6	14830,12	102820,78	250103,00	10,79
7	14672,35	101726,95	247442,33	10,71
8	14514,58	100633,12	244781,66	10,63
9	14356,81	99537,29	242120,99	10,56
10	14199,05	98445,45	239460,32	10,48

Таблица 2. Коэффициенты регрессии по себестоимости и данные по их оценке

Вид (коэффициента)	Коэффициент регрессии	Стандартная ошибка коэффициента	Уровень значимости нулевой гипотезы (P -значение)
Нулевой	3,663	0,06802	1,037E-8
Достоверность	0,3924	0,5402	0,5162
Полнота	0,0002067	0,0005942	0,7315
Своевременность	-0,02332	0,03202	0,5174

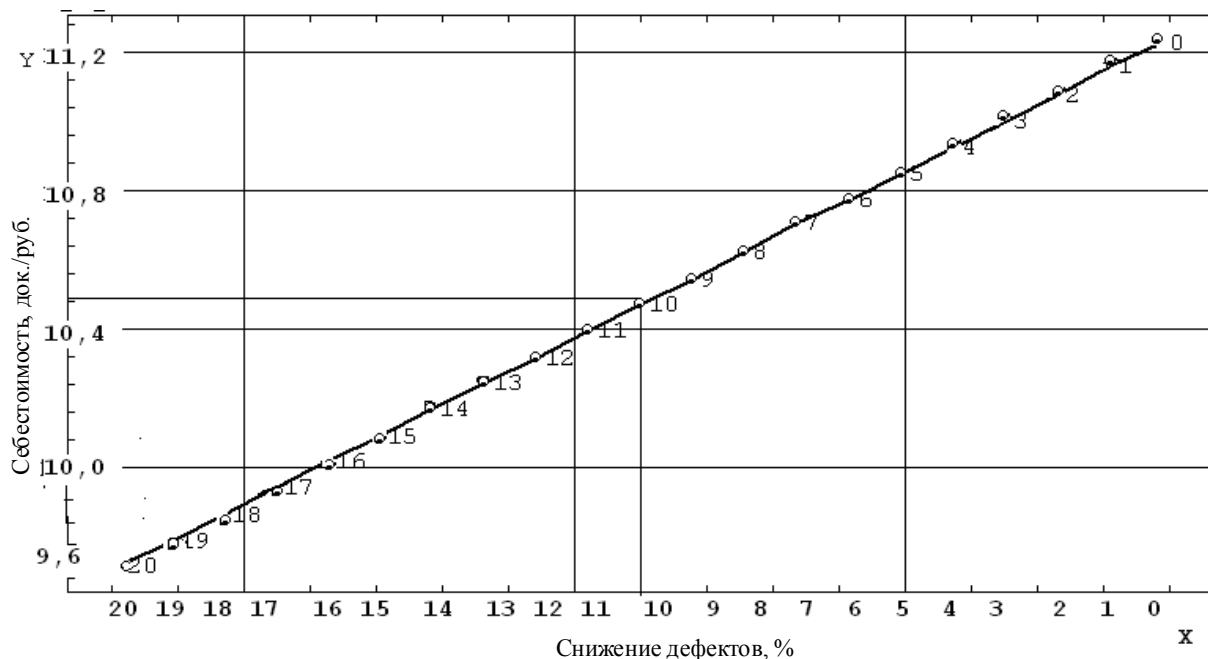


Рис. 1. График зависимости себестоимости документов от снижения трудоемкости устранения дефектов технологии ИС

(нулевой коэффициент), затем коэффициенты регрессии. Нулевой коэффициент выражает значение базового показателя себестоимости, т.е. потенциальную возможность достижения значения себестоимости выходных документов ИС при условии полного (100 %-ного) устранения дефектов - 3,66 руб. за 1 документ. Остальные коэффициенты выражают значение “весомости” переменных - дефектов по категориям, соответственно, достоверности, полноты и своевременности. Подставляя значения независимых показателей, можно определить прогнозируемые зависимые значения показателей по себестоимости.

В практических задачах оперативного управления качеством ИС на основе графика можно получать экспресс-оценки прогнозируемой вели-

чины снижения (увеличения) себестоимости. Например, необходимо определить прогнозируемое значение себестоимости документов при условии снижения дефектов на 15 %, это значение будет равно ориентировочно 10,09 руб./док (рис. 2), что вполне согласуется с данными табл. 1. Определяется путем нанесения линий параллельно осям координат до пересечения с линией регрессии.

График показывает, что при условии снижения дефектов на 0 % себестоимость документов равна 11,24 руб./док., что также согласуется с данными табл. 1. При сопоставлении значений себестоимости при 100%-ном и при 0%-ном устранении дефектов получаем относительный уровень себестоимости $3,66/11,24 \approx 0,33$. Это значит, что ресурсы по снижению себестоимос-

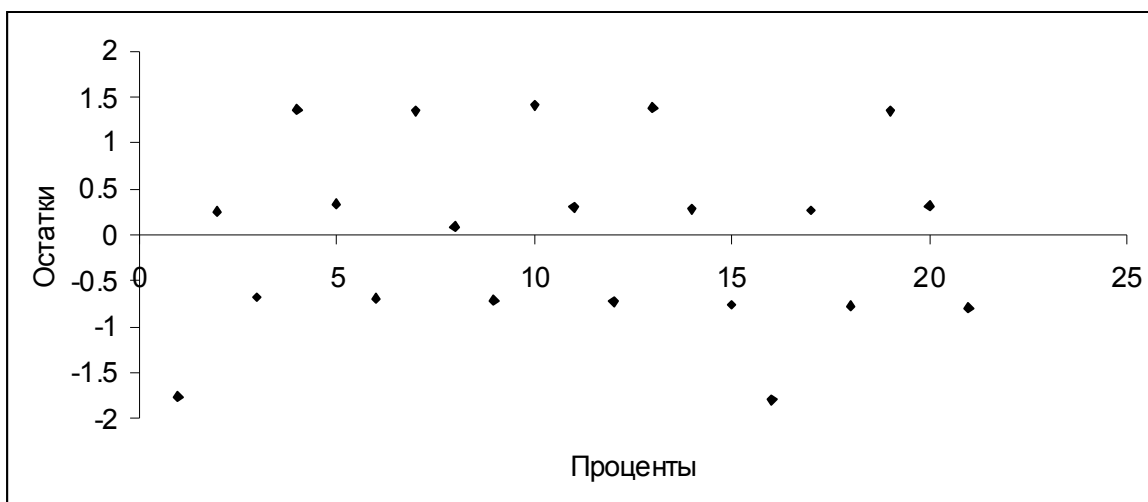


Рис. 2. График зависимости величины нормированного остатка от величины процента снижения дефектов по стоимости

Таблица 3. Данные анализа остатков регрессии по себестоимости

Значение, %	Номинальное значение себестоимости $Y_{экс}$	Значение прогноза себестоимости $Y_{рег}$	Остаток $Y_{экс} - Y_{рег}$	Остаток, ед. стандартного отклонения	Стандартная ошибка среднего значения $d/Y_{рег}$	95 % доверительного интервала $iY_{рег}$
1	2	3	4	5	6	7
0	11,24	11,25	-0,005596	-1,762	0,003446	0,007186
1	11,17	11,17	0,0007875	0,2479	0,00341	0,00711
2	11,09	11,09	-0,002168	-0,6827	0,003377	0,007042
3	11,02	11,02	0,00435	1,37	0,003348	0,006981
4	10,94	10,94	0,001064	0,3349	0,003322	0,006928
5	10,86	10,86	-0,002223	-0,6998	0,003301	0,006882
6	10,79	10,79	0,004295	1,352	0,003283	0,006845
7	10,71	10,71	0,0002629	0,08279	0,003269	0,006816
8	10,63	10,63	-0,002277	-0,717	0,003259	0,006795
9	10,56	10,56	0,004478	1,41	0,003253	0,006782
10	10,48	10,48	0,0009548	0,3006	0,003251	0,006778

ти используются только на 1/3, или имеется возможность снизить себестоимость документов на 2/3.

Для проверки адекватности модели обратимся к содержанию оценок, полученных в результате экспериментальной обработки (табл. 2). Проверка модели показала в целом по гипотезе 1 их адекватность экспериментальным данным. Относительно коэффициентов множественной корреляции, коэффициентов детерминации, критериев Фишера, уровней значимости нулевой гипотезы и других оценок можно принять, что качество линейного прогноза хорошее. На рис. 1 значения себестоимости достаточно плотно укладываются по линии регрессии, что свидетельствует о надежности модели.

Заключение можно дать на основе анализа остатков (см. табл. 3).

Значения себестоимости и анализ параметров показывает, что значения остатков незначительны. Визуальную проверку адекватности можно выполнить по графикам зависимости величин нормированных остатков от величин процентов снижения дефектов (рис. 2). График зависимости величины нормированного остатка от величины процента снижения дефектов по сто-

имости показывает отсутствие четко выраженного криволинейного тренда. Наблюдается случайный разброс, свидетельствующий о том, что модель едва ли можно или целесообразно улучшить. Сравнительная смещенность, например распределения, настолько мала (близость к оси), что не имеет принципиального значения относительно номинальных величин прогнозируемой переменной по себестоимости (см. табл. 2).

¹ Мазур И.И., Шапиро В.Д., Олдерогге Н.Г. Управление проектами. М., 2004.

² Исаев Г.Н. Управление качеством информационных систем: Теоретико-методологические основания. М., 2011.

³ Азгальдов Г.Г. Теория и практика оценки качества товаров. М., 1989.

⁴ Айвазян С.А. Прикладная статистика. Основы эконометрики. Т. 2. М., 2001.

⁵ Исаев Г.Н. Устройство для определения значений показателей качества информационных систем: патент RU \square 46371 U1 // Бюл. Федеральной службы по интеллектуальной собственности, патентам и товарным знакам. 2005. 27 июня (\square 18).

⁶ Кулаицев А.П. Методы и средства комплексного анализа данных. 4-е изд., перераб. и доп. М., 2006.

Поступила в редакцию 03.03.2012 г.