УДК 330.4

*О.В. Павлов**

ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ КРИВЫХ ОБУЧЕНИЯ ДЛЯ ПРОМЫШЛЕННЫХ ПРЕДПРИЯТИЙ

Рассматривается задача построения регрессионных моделей кривых обучения по статистическим данным предприятия АО «Салют» о производстве новых изделий. С помощью нелинейного метода наименьших квадратов определяются параметры пяти моделей кривых обучения заданного вида: Райта, Стэнфорда – В, Дейонга, логистической и экспоненциальной. В результате анализа статистических характеристик построенных моделей выбирается регрессионная зависимость, наилучшим образом аппроксимирующая данные.

Ключевые слова: кривые обучения, регрессионные модели, освоение новой продукции.

Введение

В проектах по освоению новой продукции на промышленных предприятиях проявляется эффект кривой обучения. Эффект кривой обучения заключается в том, что затраты времени работников на выполнение многократно повторяющихся производственных задач снижаются. Впервые эффект кривой обучения был описан инженером Т. Райтом в авиастроительной отрасли [1]. При каждом удвоении кумулятивного объема производства затраты времени на его производство снижаются на 10–20 процентов. Под кумулятивным объемом производства понимается количество изделий, изготовленных с начала производства продукции нарастающим итогом.

Большое количество научных публикаций, в основном иностранных, посвящено исследованию феномена кривой обучения и построению различных моделей, количественно описывающих снижение времени и затрат на выполнение производственных операций с увеличением кумулятивного объема производства.

Наиболее полно обзор, обсуждение и сравнение различных моделей кривых обучения представлены в научных публикациях [2–4].

Целью данной работы является построение регрессионных (эконометрических) моделей кривых обучения на основе статистических данных АО «Салют» об изменении трудоемкости изделий «Отсек» и «Балка» с увеличением кумулятивного объема производства. Построенные регрессионные модели имеют большое практическое значение для принятия управленческих решений о производственной деятельности предприятия. Модели кривых обучения используются для постановки различных динамических оптимизационных и игровых задач управления производственной деятельностью предприятия.

Постановка задачи и метод решения

По статистическим данным АО «Салют» о динамике изменения трудоемкости изделий «Отсек» и «Балка», с увеличением кумулятивного объема производства необходимо найти наилучшее уравнение приближенной регрессии из нескольких моделей, оценить допускаемую при этом ошибку, статистическую значимость и надежность модели.

Математически задача формулируется следующим образом. Необходимо найти уравнение регрессии заданного вида:

$$\hat{y}_x = f(x),$$

где $\hat{y}_x$ – трудоемкость изделия, теоретическое значение зависимой переменной, x – кумулятивный объем производства, независимая переменная.

Для решения задачи применяются методы регрессионного и корреляционного анализа [5]. Построение уравнения регрессии заданного вида сводится к определению его параметров. Для нахождения

Павлов Олег Валерьевич (pavlov@ssau.ru), заместитель директора, Институт экономики и управления, Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Российская Федерация, г. Самара, Московское шоссе, 34.

параметров регрессионной зависимости $f(x)$ применяется нелинейный метод наименьших квадратов [4–5]:

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_{xi}\right) \to \min, \tag{1}$$

где y_i – фактические значения трудоемкости в i-м наблюдении, $\hat{y}_{xi}$ –расчетные (теоретические) значения трудоемкости, рассчитанные по уравнению регрессии, n – число наблюдений.

Метод наименьших квадратов заключается в поиске оценок параметров уравнения регрессии, минимизирующих сумму квадратов отклонений фактических значений y_i от расчетных значений $\hat{y}_{xi}$. Рассматриваемые в работе уравнения регрессии внутренне нелинейны по оцениваемым параметрам, поэтому для решения оптимизационной задачи (1) используется численный итеративный метод Ньютона.

Для выбора наилучшей регрессионной зависимости кривой обучения рассматриваются пять различных моделей, описанных в научной литературе [1–4].

1. Модель Райта:

$$y = C_1 x^{-b}, \tag{2}$$

где y – трудоемкость изделия, C_1 – трудоемкость производства первого изделия, x – кумулятивный объем производства, b – индекс обучения, характеризует скорость снижения трудоемкости изделия при увеличении кумулятивного объема производства.

2. Модель Стэнфорда – B, которая является модификацией модели Райта:

$$y = C_1 (x + B)^{-b}, \tag{3}$$

где B – параметр, соответствующий количеству изделий предшествующего производства.

3. Модель Дейонга:

$$y = C_1\left[M + (1 - M)x^{-b}\right], \tag{4}$$

где M – коэффициент несжимаемости, который учитывает, что часть работы, выполняется машинами. Если $M = 0$, то в работе не задействовано автоматизированное оборудование, а если $M = 0$, то работа полностью выполняется автоматом и обучения не происходит.

4. Логистическая модель:

$$y = y_{\min} + (y_{\max} - y_{\min})\left[\frac{1}{1 + \beta e^{\alpha x}}\right], \tag{5}$$

где $y_{\min}$, $y_{\max}$ – минимальные и максимальные значения трудоемкости изделия, α, β – параметры логистической модели.

5. Экспоненциальная модель:

$$y = C_0 + \beta e^{-\alpha x}, \tag{6}$$

где C_0, α, β – параметры экспоненциальной модели.

Выбор наилучшей модели из построенных моделей кривых обучения осуществляется на основе расчета остаточной дисперсии $\sigma^2_{ост}$:

$$\sigma^2_{ост} = \sum_{i=1}^{n}\left(y_i - \hat{y}_{xi}\right). \tag{7}$$

Для оценки тесноты связи между зависимой и независимой переменными в построенных регрессионных моделях рассчитывается индекс парной корреляции:

$$R_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}, \tag{8}$$

где $\bar{y}$ – средняя величина зависимой переменной (трудоемкостью), определяемая по формуле

$$\bar{y} = \sum_{i=1}^{n} y_i .$$

Чем ближе индекс парной корреляции к 1, тем более тесная связь между зависимой переменной (трудоемкостью) и независимой переменной (кумулятивным объемом производства изделия).

Для оценки качества построенных регрессионных моделей вычисляется индекс детерминации как квадрат индекса парной корреляции:

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

Индекс детерминации показывает долю изменения зависимой переменной (трудоемкости), обусловленную изменением независимой переменной (кумулятивным объемом производства). Чем ближе индекс детерминации к 1, тем лучше качество построенной регрессионной модели.

Для построенных регрессионных моделей проводится оценка их значимости с помощью F-критерия Фишера. При этом выдвигается гипотеза о статистической незначимости уравнения регрессии и показателя тесноты связи. Выполняется сравнение фактического $F_{факт}$ и табличного $F_{табл}$ значений F-критерия.

Величина фактического F-критерия связана с индексом детерминации:

$$F_{факт} = \frac{R^2}{1 - R^2} \frac{n - m - 1}{m}, \tag{10}$$

где m – число параметров при независимой переменной x.

Табличное значение $F_{табл}$ определяется по таблице критических значений при уровне значимости α и двух степенях свободы $k_1 = m$, $k_2 = n - m - 1$. В расчетах уровень значимости α принимается равным 0,01.

Если фактическое значение больше табличного, то гипотеза о случайной природе оцениваемых переменных отклоняется и признается статистическая значимость и надежность уравнения регрессии. В противном случае гипотеза не отклоняется и признается статистическая незначимость, ненадежность построенной регрессионной модели.

Качество анализируемых нелинейных уравнений регрессии также оценивается по средней ошибке аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{xi})}{y_i} 100\% . \tag{11}$$

Значение средней ошибки аппроксимации до 10 % свидетельствует о высокой точности регрессионной модели.

Результаты построения регрессионной модели кривой обучения по данным производства изделия «Отсек»

Статистические данные о динамике трудоемкости изделия «Отсек» от кумулятивного объема производства на предприятии ОА «Салют» представлены на рис. 1.

На графике рис. 1 видно, что точки выстраиваются в кривую линию, соответствующую некоторому нелинейному регрессионному уравнению *f(x)*. В результате решения оптимизационной задачи (1) с использованием инструмента «Поиск решения электронной таблицы Excel» были определены параметры пяти моделей кривых обучения (2)–(6) по статистическим данным производства изделия «Отсек».

1. Модель Райта: $y = 205{,}45x^{-0{,}2510}$.
2. Модель Стэнфорда – В: $y = 266{,}87(x + 10)^{-0{,}3005}$.
3. Модель Дейонга: $y = 238{,}76(0{,}10 + 0{,}9x^{-0{,}3617})$.
4. Логистическая модель: $y = 55{,}10 + 36{,}61\left[\frac{1}{1 + 0{,}017e^{0{,}0561x}}\right]$.
5. Экспоненциальная модель: $y = 49{,}85 + 59{,}51e^{-0{,}0138x}$.

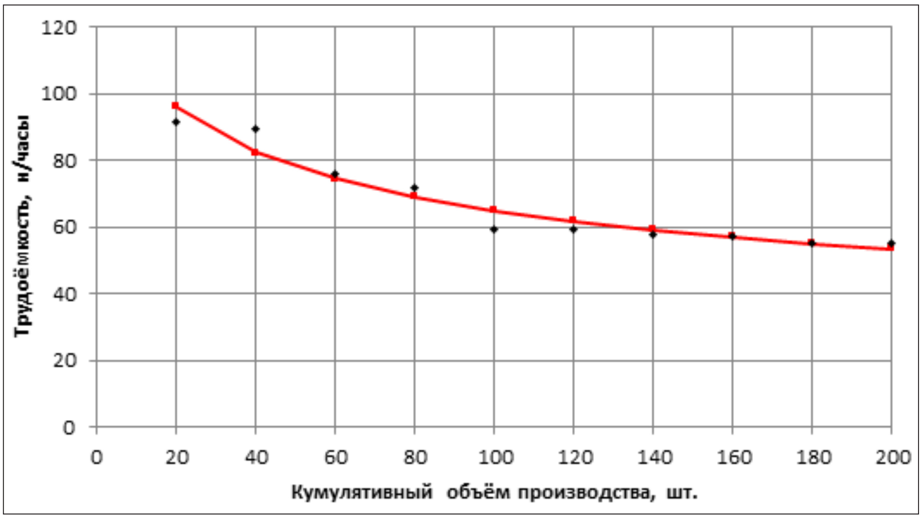

Рис. 1. Зависимость трудоемкости изделия «Отсек» от кумулятивного объема производства

Построенные регрессионные модели по статистическим данным производства изделия «Отсек» представлены на рис. 2–6.

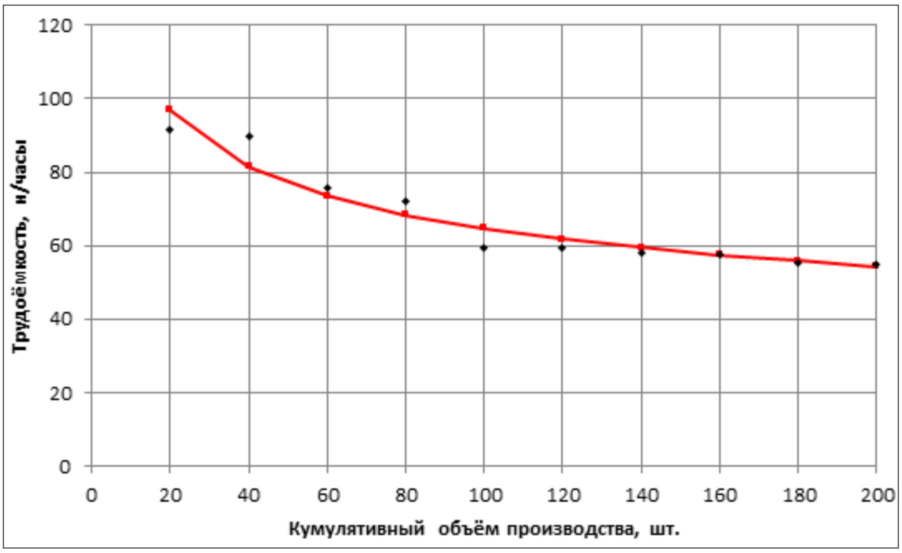

Рис. 2. Регрессионная модель Райта кривой обучения по данным производства изделия «Отсек»

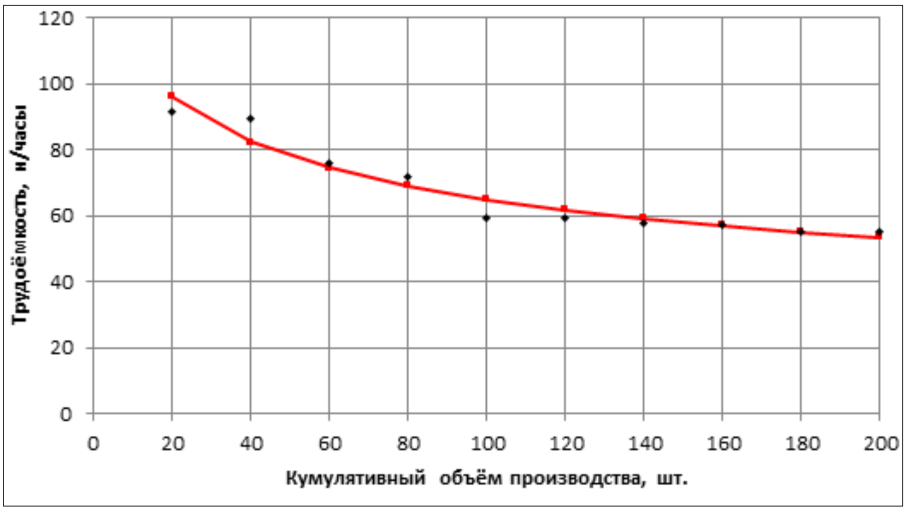

Рис. 3. Регрессионная модель Стэнфорда – В кривой обучения по данным производства изделия «Отсек»

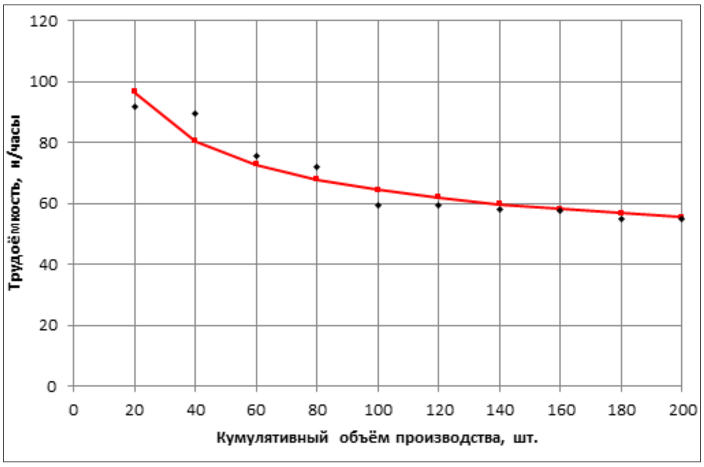

Рис. 4. Регрессионная модель Дейонга кривой обучения
по данным производства изделия «Отсек»

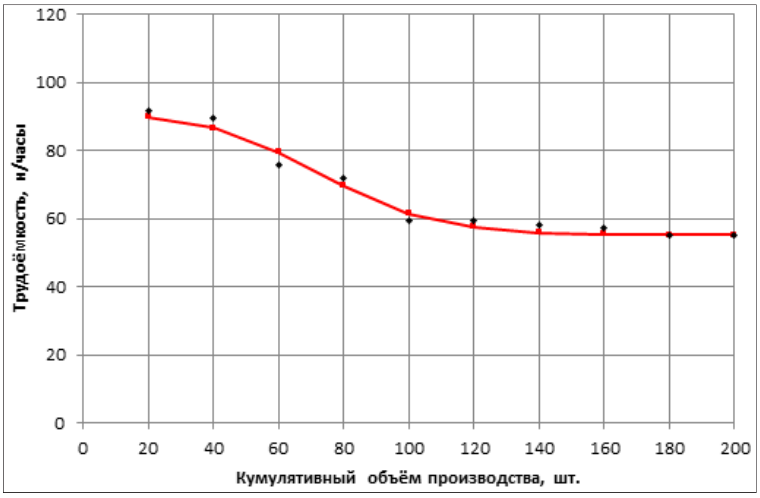

Рис. 5. Регрессионная логистическая модель кривой обучения
по данным производства изделия «Отсек»

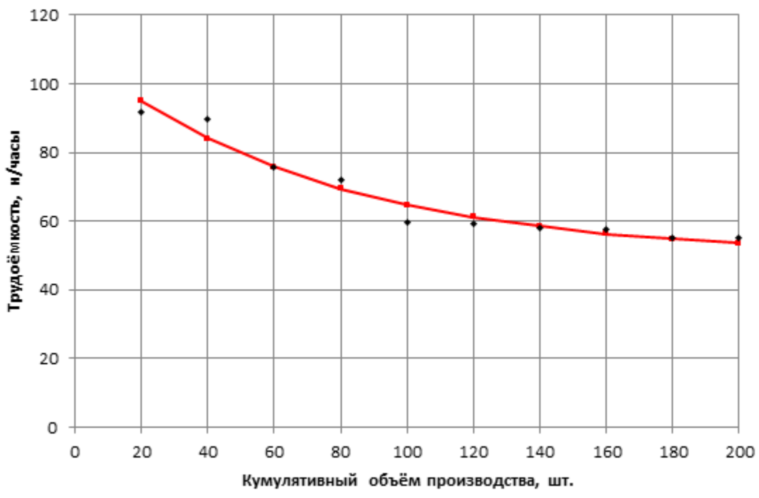

Рис. 6. Регрессионная экспоненциальная модель кривой обучения
по данным производства изделия «Отсек»

В таблице 1 для построенных моделей по данным производства изделия «Отсек» приводятся рассчитанные по формулам (7)–(11) значения остаточной дисперсии $\sigma^2_{ост}$, индекса парной корреляции R_{xy}, индекса детерминации R^2, фактического значения F-критерия Фишера $F_{факт}$, средней ошибки аппроксимации $\overline{A}$.

Таблица 1

Расчет статистических характеристик моделей кривых обучения по данным производства изделия «Отсек»

№	Модель	Остаточная дисперсия	Индекс парной корреляции	Индекс детерминации	Фактическое значение F-критерия Фишера	Средняя ошибка аппроксимации
1	Райта	14,7890	0,9579	0,9175	89,0187	4,08%
2	Стэнфорда – В	12,2224	0,9653	0,9319	109,3917	3,79%
3	Дейонга	17,1486	0,9510	0,9044	75,6691	4,61%
4	**Логистическая**	**4,8268**	**0,9865**	**0,9731**	**289,2602**	**2,74%**
5	Экспоненциальная	8,2415	0,9768	0,9540	166,0952	3,12%

Анализируя данные таблицы 1, приходим к выводу, что у логистической модели по сравнению с другими моделями остаточная дисперсия $\sigma^2_{ост}$ и средняя ошибка $\overline{A}$ наименьшие, индекс парной корреляции R_{xy}, индекс детерминации R_{xy} наибольшие. Таким образом, наиболее хорошо статистические данные по производству изделия «Отсек» аппроксимирует логистическая модель.

Для оценки значимости построенной логистической модели было определено табличное значение F-критерия Фишера $F_{табл} = 11{,}2586$. Так как фактическое значение больше табличного значения F-критерия Фишера $F_{факт} > F_{табл}$, то гипотеза о случайной природе оцениваемых переменных отклоняется и признается статистическая значимость и надежность экспоненциальной модели.

Значение средней ошибки аппроксимации $\overline{A} = 2{,}74$ % , меньшее 10 %, свидетельствует о высокой точности построенной логистической модели.

Результаты построения регрессионной модели кривой обучения по данным производства изделия «Балка»

Статистические данные зависимости трудоемкости изделия «Балка» от кумулятивного объема производства на предприятии АО «Салют» представлены на рис. 7.

Из анализа рис. 7 видно, что точки выстраиваются в кривую линию, соответствующую некоторому нелинейному регрессионному уравнению *f(x)*. В результате решения оптимизационной задачи (1) с использованием инструмента «Поиск решения электронной таблицы Excel» были определены параметры пяти моделей кривых обучения (2)–(6) по статистическим данным производства изделия «Балка».

1. Модель Райта: $y = 20{,}68x^{-0{,}1469}$.
2. Модель Стэнфорда – В: $y = 22{,}57(x + 5)^{-0{,}1630}$.
3. Модель Дейонга: $y = 21{,}38(0{,}1 + 0{,}9x^{-0{,}1813})$.
4. Логистическая модель: $y = 9{,}10 + 5{,}20\left[\dfrac{1}{1+0{,}119e^{0{,}0341x}}\right]$.
5. Экспоненциальная модель: $y = 9{,}17 + 6{,}16e^{-0{,}0169x}$.

Построенные регрессионные модели по статистическим данным производства изделия «Балка» представлены на рис. 8–12.

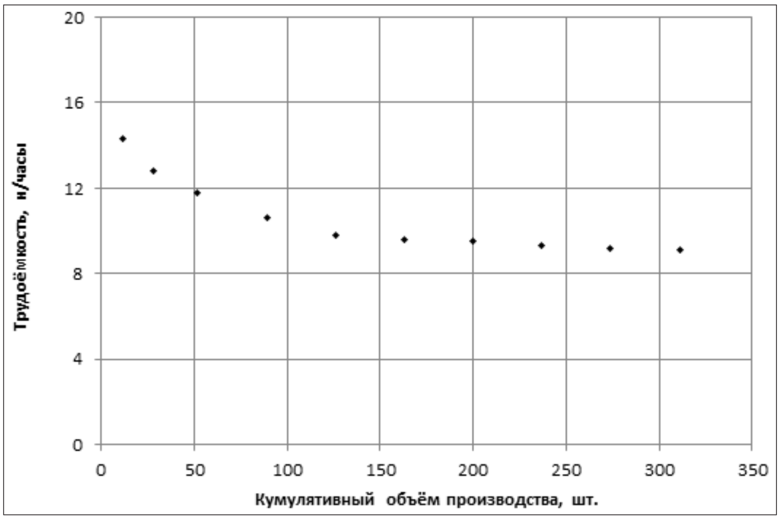

Рис. 7. Зависимость трудоемкости изделия «Балка» от кумулятивного объема производства

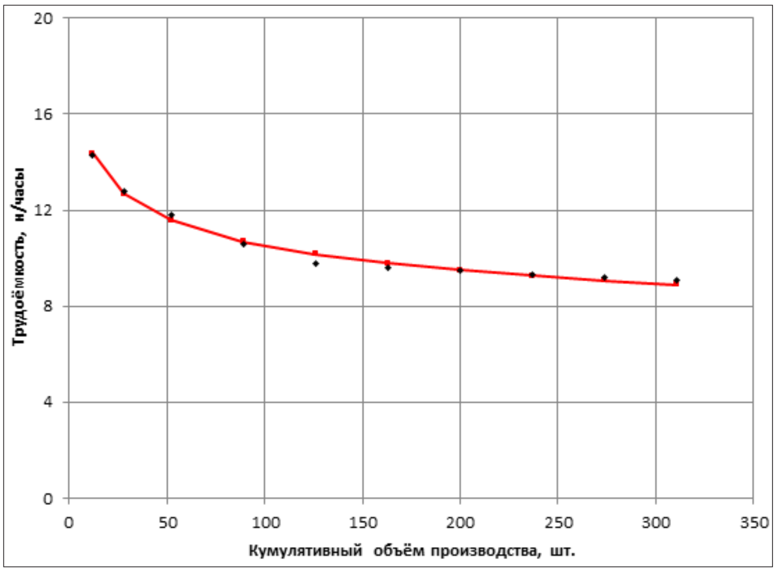

Рис. 8. Регрессионная модель Райта кривой обучения по данным производства изделия «Балка»

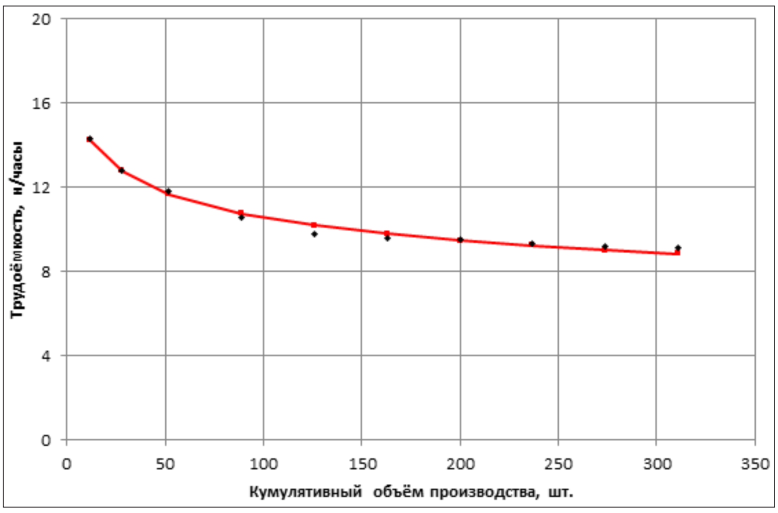

Рис. 9. Регрессионная модель Стэнфорда – В кривой обучения по данным производства изделия «Балка»

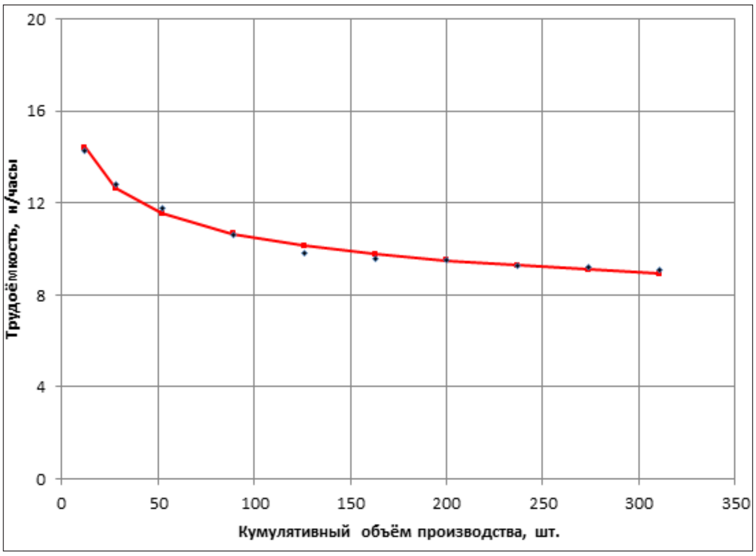

Рис. 10. Регрессионная модель Дейонга кривой обучения по данным производства изделия «Балка»

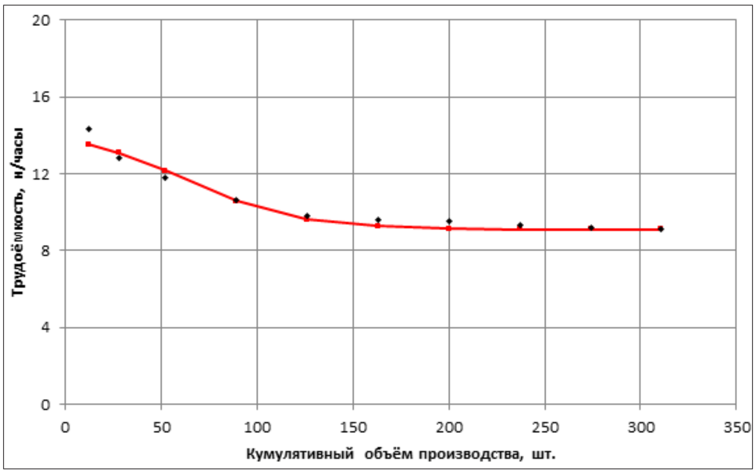

Рис. 11. Регрессионная логистическая модель кривой обучения по данным производства изделия «Балка»

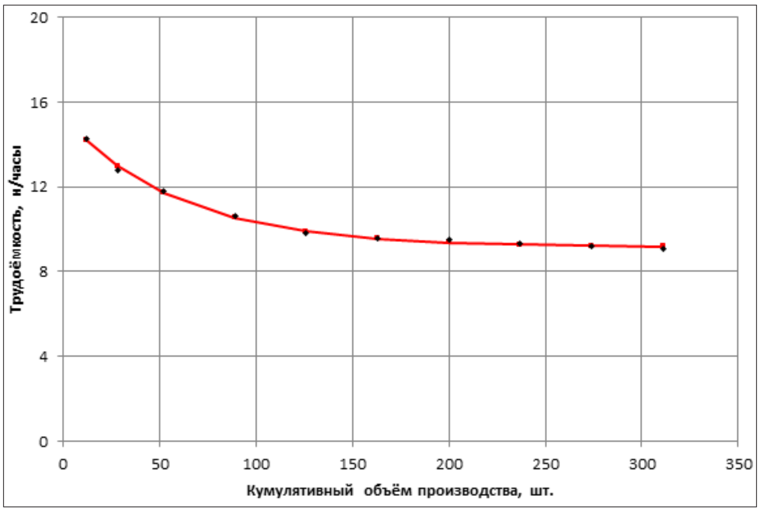

Рис. 12. Регрессионная экспоненциальная модель кривой обучения по данным производства изделия «Балка»

В таблице 2 для построенных моделей по данным производства изделия «Балка» приводятся рассчитанные по формулам (7)–(11) значения остаточной дисперсии $\sigma^2_{ост}$, индекса парной корреляции R_{xy}, индекса детерминации R^2, фактического значения F-критерия Фишера $F_{факт}$, средней ошибки аппроксимации $\bar{A}$.

Таблица 2

Расчет статистических характеристик моделей кривой обучения по данным производства изделия «Балка»

№	Модель	Остаточная дисперсия	Индекс парной кореляции	Индекс детерминации	Фактическое значение F-критерия Фишера	Средняя ошибка аппроксимации
1	Райта	0,0304	0,9947	0,9894	747,3331	1,39%
2	Стэнфорда-В	0,0354	0,9938	0,9877	640,9246	1,53%
3	Дейонга	0,0294	0,9949	0,9898	772,5261	1,33%
4	Логистическая	0,1135	0,9800	0,9605	194,3788	2,27%
5	**Экспоненциальная**	**0,0100**	**0,9983**	**0,9965**	**2286,3202**	**0,79%**

Анализируя данные таблицы 2, приходим к выводу, что у экспоненциальной модели по сравнению с другими моделями остаточная дисперсия $\sigma^2_{ост}$ и средняя ошибка $\bar{A}$ наименьшие, индекс парной корреляции R_{xy}, индекс детерминации R_{xy} наибольшие. Таким образом, наиболее хорошо статистические данные по производству изделия «Балка» аппроксимирует экспоненциальная модель.

Для оценки значимости построенной экспоненциальной модели определяется табличное значение F-критерия Фишера $F_{табл}$ 11,2586. Так как фактическое значение больше табличного значения F-критерия Фишера $F_{факт} > F_{табл}$, то гипотеза о случайной природе оцениваемых переменных отклоняется и признается статистическая значимость и надежность экспоненциальной модели.

Значение средней ошибки аппроксимации $\bar{A}$ 0,79 %, меньшее 10 %, свидетельствует о высокой точности построенной экспоненциальной модели.

Заключение

В данной работе на основе статистических данных предприятия АО «Салют» о производстве новых изделий «Отсек» и «Балка» построены регрессионные (эконометрические) модели кривых обучения. Для имеющихся статистических данных построены пять моделей: Райта, Стэнфорда–В, Дейонга, логистическая и экспоненциальная. Для каждой модели были рассчитаны статистические характеристики. Из анализа графиков регрессионных уравнений и рассчитанных статистических характеристик моделей были сделаны следующие выводы:

1. Наиболее хорошо статистические данные по производству изделия «Отсек» аппроксимирует логистическая модель следующего вида:

$$y = 55{,}10 = 36{,}61\left[\frac{1}{1+0{,}017e^{0{,}0561x}}\right].$$

2. Наиболее хорошо статистические данные по производству изделия «Балка» аппроксимирует экспоненциальная модель следующего вида:

$$y = 9{,}17 + 6{,}16e^{-0{,}0169x}.$$

3. Проведенное исследование показало статистическую значимость и надежность выбранных логистической и экспоненциальной моделей.

4. Расчеты свидетельствуют о высокой точности построенных моделей.

Построенные регрессионные модели кривых обучения будут использованы автором для постановки различных динамических оптимизационных и игровых задач управления производственной деятельностью предприятий.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ и Самарской области в рамках научного проекта № 17-46-630606.

Библиографический список

1. Wright T.P. Factors affecting the cost of airplanes // Journal of the aeronautical sciences. 1936. V. 3.№ 4. P. 122–128.
2. Badiru A. Computational survey of univariate and multivariate learning curvemodels // IEEE Transactions on Engineering Management. 1992. V. 39. № 2. P. 176–188.
3. Yelle L.E. The learning curve: Historical review and comprehensive survey // Decision Sciences. 1979. V. 10. № 2. P. 302–328.
4. Learning Curves: Theory, Models, and Applications / ed. by Mohamad Y. Jaber. Boca Raton: CRC Press, 2011. 476 p.
5. Эконометрика: учебник / под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. 344 с.

References

1. Wright T.P. Factors affecting the cost of airplanes. In: Journal of the aeronautical sciences, 1936, vol. 3, no. 4, pp. 122–128 [in English].
2. Badiru A. Computational survey of univariate and multivariate learning curvemodels. In: IEEE Transactions on Engineering Management, 1992, vol. 39, no. 2, pp. 176–188 [in English].
3. Yelle L.E. The learning curve: Historical review and comprehensive survey. In: Decision Sciences, 1979, vol. 10, no. 2, pp. 302–328 [in English].
4. Learning Curves: Theory, Models, and Applications.Edited by Mohamad Y. Jaber. Boca Raton: CRC Press, 2011, 476 p. [in English].
5. Ekonometrika: uchebnik [Econometrics: a textbook]. Ed. by I.I. Yeliseyeva. Moscow: Finansy I statistika, 2003. 344 p. [in Russian].

O.V. Pavlov*

CONSTRUCTION OF REGRESSION MODELS OF LEARNING CURVES FOR THE INDUSTRIAL ENTERPRISES

The problem of constructing regression models of learning curves is considered based on the statistical data of the Salyut JSC enterprise on the production of new products.Using the nonlinear method of least squares, the parameters of five models of learning curves of a given type are defined: Wright's, Stanford-B, Deyong's, logistic and exponential.As a result of the statistical characteristics analysis of the constructed models, the regression dependence is chosen that best approximates the data.

Key words: learning curves, regression models, new products development.

Статья поступила в редакцию 4/IX/2017.
The article received 4/IX/2017.

* *Pavlov Oleg Valerievich* (pavlov@ssau.ru), Vice Director, Institute of Economics and Management, Samara National Research University, 34, Moskovskoye shosse, Samara, 443086, Russian Federation.