

УДК 622.276.432:004.9

## ПРИМЕНЕНИЕ ФАКТОРНОГО АНАЛИЗА ДЛЯ РЕДУКЦИИ ПРИЗНАКОВОГО ПРОСТРАНСТВА НАГНЕТАТЕЛЬНЫХ СКВАЖИН

**А. Ю. Вирстюк<sup>1</sup>, В. С. Микшина<sup>2</sup>**

<sup>1</sup> Сургутский государственный университет, Nastua5.1991@mail.ru

<sup>2</sup> Сургутский государственный университет, Mikshinavs@gmail.ru

В данной статье рассмотрена проблема определения эффективности работы нефтяных скважин. Проведены выбор и краткое описание показателей работы нагнетательных скважин, оказывающих косвенное влияние на основную характеристику нефтяной скважины – дебит. Между выделенными характеристиками нагнетательных скважин установлена взаимозависимость, в связи с чем проведен факторный анализ с целью объединения их в группы (факторы), независимые между собой. Для определения и прогнозирования эффективности работы нефтяных скважин между полученными факторами и дебитом установлена зависимость в виде уравнения регрессии. Для проведения многомерного анализа исходных данных применен программный пакет статистической обработки данных Statistica.

*Ключевые слова:* факторный анализ, нагнетательная скважина, нефтяная скважина.

## FACTOR ANALYSIS APPLICATION FOR REDUCTION OF THE ATTRIBUTE SPACE OF INJECTION WELLS

**A. Yu. Virstyuk<sup>1</sup>, V. S. Mikshina<sup>2</sup>**

<sup>1</sup> Surgut State University, Nastua5.1991@mail.ru

<sup>2</sup> Surgut State University, Mikshinavs@gmail.ru

The article considers the problem of determining the efficiency of oil wells. A choice and a brief description of performance indicators of injection wells are presented. These indicators have the indirect effect on the main feature of an oil well – the production rate. The correlation between highlighted characteristics is established; therefore the factor analysis to combine them into the independent groups (factors) is carried out. The regression equation for determining and predicting the efficiency of wells is applied to describe the relation between obtained factors and the production rate. Software package STATISTICA is used for the multivariate analysis of the initial data.

*Keywords:* factor analysis, injection well, oil well.

Окружающий нас мир многомерен в том смысле, что каждый объект характеризуется множеством параметров. Человеческий мозг не способен одновременно оперировать более чем тремя измерениями пространства, что приводит к сложностям при анализе данных. Для сведения многомерных данных к нескольким измерениям с минимальными потерями информации была разработана специальная группа методов статистического анализа данных – многомерный анализ данных.

Одним из методов статистического многомерного анализа является факторный анализ. Этот метод применяется при наличии взаимозависимых показателей и преследует две цели: сокращение числа переменных и классификация переменных с целью определения структуры взаимосвязей между ними [6].

В рамках данной работы будет рассмотрено применение факторного анализа для редукции признакового пространства нагнетательных скважин с целью дальнейшего прогнозирования дебита нефтяных скважин и оценки их эффективности.

Эффективность работы нефтяных скважин во многом определяется такими показателями, как: вязкость, нефтенасыщенность, однородность, темпы отбора жидкости и пр., то есть группами геологических и технологических параметров, число которых может измеряться несколькими десятками.

Но, зачастую, большой, хотя и косвенный вклад в объемы добываемой нефти вносят нагнетательные скважины, которые оказывают влияние на зависимые от них нефтяные скважины, осуществляя закачку воды в пласты. В геологии данный процесс носит название интерференции.

Поэтому очень важно при оценке работы нефтяных скважин учитывать и параметры влияющих на них нагнетательных скважин. Для того, чтобы исключить взаимозависимость между показателями работы нагнетательных скважин и объединить наиболее близкие из них в группы (так называемые факторы), требуется применение факторного анализа.

В общем виде, зависимость между дебитом нефтяной скважины и нормированными показателями работы влияющей нагнетательной скважины, объединенными в группы (факторы), можно представить формулой (1):

$$Y = b_0 + b_1 F_1 + b_2 F_2 + \dots + b_m, \quad (1)$$

где  $Y$  – дебит нефтяной скважины;

$b_0$  – прочие факторы, влияющие на работу нефтяной скважины;

$b_i$  – коэффициенты уравнения регрессии;

$F_i$  – группы (факторы) показателей работы влияющих нагнетательных скважин.

Наиболее важные показатели работы нагнетательных скважин, к которым и будет применен факторный анализ, представлены ниже:

1. Время работы скважины (час) –  $P_1$
2. Объем закачиваемой жидкости ( $\text{м}^3$ ) –  $P_2$
3. Давление закачки или буферное давление (ат) –  $P_3$
4. Приемистость ( $\text{м}^3/\text{сут}$ ) –  $P_4$
5. Коэффициент охвата пласта заводнением (%) –  $P_5$
6. Эффективная мощность пласта (м) –  $P_6$ .

Если первые три показателя интуитивно понятны, то остальные следует рассмотреть подробнее.

Приемистость нагнетательной скважины определяется объемом воды (жидкости, реагентов), закачиваемым в пласт в единицу времени. Коэффициент охвата пласта заводнением рассчитывается при анализе профиля приемистости скважины по формуле (2):

$$K_{\text{ИП}}^{\text{охв зав}} = \left( \frac{H_1^{\text{раб}} + H_2^{\text{раб}} + \dots + H_n^{\text{раб}}}{\sum_1^i H_i^{\text{раб+нераб}}} \right) \times 100, \quad (2)$$

где  $K_{\text{ИП}}^{\text{охв зав}}$  – охват пласта заводнением в интервале перфорации, %;

$H_1^{\text{раб}}, H_2^{\text{раб}}, \dots, H_n^{\text{раб}}$  – мощность 1, 2..... $n$  работающих пропластков в интервале перфорации, м;

$\sum_1^i H_i^{\text{раб+нераб}}$  – суммарная мощность работающих и неработающих пропластков в интервале перфорации, м [1–2].

Эффективная мощность пласта представляет собой суммарную мощность (толщину) нефтяных (газонасыщенных) прослоев в пласте, способных отдавать нефть (газ).

Для проведения факторного анализа были отобраны данные со 100 нагнетательных скважин. Полученные данные актуальны на декабрь 2017 г. Так как исходные показатели несопоставимы в силу различной размерности, то сначала целесообразно получить нормированные (стандартизированные) данные, которые рассчитываются по формуле (3):

$$X_{i,j} = \frac{P_{i,j} - \mu_i}{\sigma_i}, \quad (3)$$

где  $X_{i,j}$  – нормированное значение показателя;

$P_{i,j}$  – исходное значение показателя;

$\mu_i$  – математическое ожидание  $i$ -ого показателя работы, полученное по 100 нагнетательным скважинам;

$\sigma_i$  – среднеквадратическое отклонение [3].

В табл. 1 представлены показатели работы по некоторым нагнетательным скважинам, а также  $\mu_i$  и  $\sigma_i$  для каждого показателя, необходимые для стандартизации исходных значений.

Таблица 1

## Показатели работы нагнетательных скважин и параметры выборки

| №          | $P_1$  | $P_2$   | $P_3$  | $P_4$ | $P_5$ | $P_6$ |
|------------|--------|---------|--------|-------|-------|-------|
| 1          | 720    | 27979   | 137    | 935   | 81,82 | 11    |
| 2          | 720    | 4316    | 144    | 150   | 100   | 8     |
| 3          | 720    | 18943   | 137    | 590   | 100   | 4     |
| 4          | 720    | 17574   | 97     | 429   | 17,95 | 4,6   |
| 5          | 720    | 1663    | 137    | 58    | 100   | 2     |
| 6          | 528    | 55      | 130    | 2     | 100   | 3,2   |
| 7          | 720    | 6635    | 123    | 130   | 100   | 3,8   |
| 8          | 720    | 3775    | 129    | 178   | 100   | 4,2   |
| 9          | 96     | 23      | 122    | 10    | 94,2  | 13,8  |
| 10         | 720    | 79      | 127    | 3     | 7,79  | 15,4  |
| ...        | ...    | ...     | ...    | ...   | ...   | ...   |
| 100        | 720    | 3775    | 127    | 450   | 100   | 4     |
| $\mu_i$    | 679,92 | 5529    | 104,01 | 184   | 85,43 | 9,57  |
| $\sigma_i$ | 140,82 | 5337,61 | 27,57  | 180   | 22,58 | 6,59  |

Как уже отмечалось ранее, показатели работы нагнетательных скважин характеризуются взаимозависимостью, исключить которую можно путем объединения их в факторы. Для оценки взаимозависимости необходимо построить корреляционную матрицу, которая представляет собой матрицу коэффициентов корреляции. Кроме того, корреляционная матрица позволяет определить возможность сокращения признаков пространства.

Связь между показателями работы скважины можно считать сильной, если коэффициент корреляции составляет от  $\pm 0,7$  до  $\pm 1$ ; средней – если коэффициент имеет значения от  $\pm 0,3$  до  $\pm 0,699$ . В остальных случаях связь можно считать слабой [7].

В табл. 2 представлена корреляционная матрица, полученная на основе нормированных показателей работы нагнетательных скважин. Из нее видно, что сильная и средняя корреляционная связь проявляется между нормированными переменными  $X_2$  и  $X_4$ ,  $X_3$  и  $X_6$ ,  $X_5$  и  $X_6$ .

Таблица 2

## Корреляционная матрица

|       | $X_1$ | $X_2$       | $X_3$        | $X_4$       | $X_5$        | $X_6$        |
|-------|-------|-------------|--------------|-------------|--------------|--------------|
| $X_1$ | 1,00  | 0,17        | -0,07        | 0,13        | 0,13         | 0,00         |
| $X_2$ | 0,17  | 1,00        | 0,01         | <b>0,89</b> | -0,15        | 0,10         |
| $X_3$ | -0,07 | 0,01        | 1,00         | 0,04        | 0,03         | <b>-0,25</b> |
| $X_4$ | 0,13  | <b>0,89</b> | 0,04         | 1,00        | -0,15        | 0,15         |
| $X_5$ | 0,13  | -0,15       | 0,03         | -0,15       | 1,00         | <b>-0,27</b> |
| $X_6$ | 0,00  | 0,10        | <b>-0,25</b> | 0,15        | <b>-0,27</b> | 1,00         |

Соответственно, предварительный анализ данных подтверждает возможность сокращения размерности признакового пространства и выделение факторов, число которых будет меньше числа исходных переменных (т. е. меньше 6). Полученные таким образом факторы будут исключать взаимозависимость исходных показателей и упростят уравнение регрессии (1), позволяющее прогнозировать эффективность работы нефтяной скважины в зависимости от показателей работы влияющих нагнетательных скважин.

Для редукции исходных данных был выбран программный пакет Statistica. В ходе проведения факторного анализа методом главных компонент каждому выделенному фактору ставится в соответствие своя дисперсия, называемая собственным значением [4–5, 8]. Если предположить, что число главных факторов равно числу исходных переменных, т. е. 6, то можно получить следующие значения:

Таблица 3

### Некоторые характеристики выделенных факторов

| № | Собственные значения | % от общей дисперсии |
|---|----------------------|----------------------|
| 1 | 2,018785             | 33,64642             |
| 2 | 1,333924             | 22,23206             |
| 3 | 1,119885             | 18,66475             |
| 4 | 0,775039             | 12,91731             |
| 5 | 0,642738             | 10,71229             |
| 6 | 0,109629             | 1,82716              |

Собственное значение фактора, меньшее единицы, указывает на то, что изменчивость соответствующего главного фактора меньше изменчивости любого из исходных показателей, и тогда такие факторы исключаются из дальнейшего анализа в силу своей малой информативности [3]. Этот метод определения числа главных факторов носит название метода Кайзера. Из табл. 3 видно, что у первых трех выделенных факторов собственные значения больше 1. Это указывает на то, что фактор выделяет дисперсию, эквивалентную, по крайней мере, дисперсии одной переменной.

Далее следует рассмотреть матрицу с факторными нагрузками, которая показывает корреляции между переменными и выделенными факторами.

Таблица 4

### Матрица факторных нагрузок

|       | $F_1$         | $F_2$         | $F_3$         | $F_4$  | $F_5$  | $F_6$  |
|-------|---------------|---------------|---------------|--------|--------|--------|
| $X_1$ | -0,237        | -0,266        | <b>0,727</b>  | 0,581  | -0,087 | -0,008 |
| $X_2$ | <b>-0,931</b> | -0,224        | -0,029        | -0,164 | -0,042 | 0,233  |
| $X_3$ | 0,059         | -0,581        | <b>-0,592</b> | 0,432  | 0,350  | 0,014  |
| $X_4$ | <b>-0,936</b> | -0,198        | -0,069        | -0,157 | 0,040  | -0,234 |
| $X_5$ | 0,333         | <b>-0,551</b> | 0,470         | -0,423 | 0,431  | 0,005  |
| $X_6$ | -0,327        | <b>0,730</b>  | 0,122         | 0,143  | 0,569  | 0,020  |

Из табл. 4 видно, что первым трем факторам соответствуют большие значения коэффициентов корреляции, чем остальным факторам. Причем, с увеличением номера фактора значения коэффициентов корреляции стремительно уменьшаются. Как известно, факторные нагрузки должны объединять переменные в группы, для которых коэффициенты корреляции с факторами принимают большие значения по одной группе и меньшие значения по другой [5, 9]. В данном случае, в первом факторе доминируют первый и третий показатели, второй описывает пятый и шестой, а третий фактор, в свою очередь, первый и третий исходный показатель.

Матрица факторных нагрузок показывает, что остальные факторы являются малоинформативными, поэтому их не стоит интерпретировать.

В основном, процедура поиска главных факторов подобна вращению, максимизирующему дисперсию (варимакс) исходного пространства переменных [3, 10]. Вращение факторов приводит к повороту осей, что, в свою очередь, позволяет получить интерпретируемое решение. К примеру, второй фактор (табл. 4), описывающий пятый и шестой исходные показатели работы скважин, имеет достаточно высокий коэффициент и для третьего параметра, что ставит под сомнение правильность выделения факторов.

После проведения процедуры *Varimax row* (варимакс исходных) пакета Statistica была получена матрица факторных нагрузок (табл. 5).

Таблица 5

### Матрица факторных нагрузок после вращения

|       | $F_1$        | $F_2$         | $F_3$         |
|-------|--------------|---------------|---------------|
| $X_1$ | 0,409        | -0,138        | <b>-0,638</b> |
| $X_2$ | <b>0,973</b> | 0,068         | 0,026         |
| $X_3$ | 0,235        | -0,117        | <b>0,777</b>  |
| $X_4$ | <b>0,965</b> | 0,045         | 0,022         |
| $X_5$ | -0,183       | <b>-0,801</b> | -0,249        |
| $X_6$ | -0,031       | <b>0,797</b>  | -0,321        |

На основе табл. 5 можно сделать вывод, что выполненный поворот осей подтверждает правильность выделения главных факторов, так как первый фактор отвечает за процессы закачки, второй – за профиль приемистости, а третий – за режим нагнетания, так как в течение всего периода работы нагнетательной скважины давление закачки (нагнетания) должно быть оптимальным.

После всех выполненных процедур можно записать линейные уравнения регрессии, которые связывают исходные показатели работы нагнетательных скважин и выделенные факторы:

$$F_1 = 0,21 * X_1 + 0,46 * X_2 + 0,11 * X_3 + 0,45 * X_4 - 0,04 * X_5 - 0,05 * X_6, \quad (4)$$

$$F_2 = -0,15 * X_1 + 0 * X_2 - 0,07 * X_3 - 0,02 * X_4 - 0,61 * X_5 + 0,6 * X_6, \quad (5)$$

$$F_3 = -0,55 * X_1 + 0,01 * X_2 + 0,65 * X_3 + 0,01 * X_4 - 0,24 * X_5 - 0,24 * X_6. \quad (6)$$

Учитывая вклад каждого показателя в фактор на основе матрицы факторных нагрузок, полученной после вращения, можно уравнения (4)–(6) свести к виду (7)–(9):

$$F_1 = 0,46 * X_2 + 0,45 * X_4, \quad (7)$$

$$F_2 = -0,61 * X_5 + 0,6 * X_6, \quad (8)$$

$$F_3 = -0,55 * X_1 + 0,65 * X_3. \quad (9)$$

Предварительный графический анализ регрессионных остатков подтвердил адекватность построенных моделей для каждого выделенного фактора, об этом свидетельствует стремление распределения к нормальному виду.

Уравнение регрессии (1), определяющее зависимость дебита интерферирующих нефтяных скважин от эффективности работы нагнетательной скважины, можно представить в следующем виде:

$$Y = b_0 + b_1 F_1 + b_2 F_2 + b_3 F_3, \quad (10)$$

где  $b_i$  – коэффициенты регрессии, которые будут получены экспериментальным путем;

$b_0$  – прочие факторы, влияющие на работу нефтяной скважины;

$F_i$  – факторы, объединяющие показатели работы нагнетательных скважин.

Таким образом, можно сделать вывод, что факторный анализ применим для исследования нагнетательных скважин с целью снижения признакового пространства. В

работе была подтверждена возможность сокращения 6 исходных признаков до 3 главных факторов. Для этого были получены и проанализированы корреляционная матрица, матрица факторных нагрузок до и после вращения, собственные значения. Выделенный первый фактор отвечает за процессы закачки, второй – за профиль приемистости нагнетательной скважины, а третий – за режим нагнетания. С их помощью составлено уравнение регрессии (10), которое в дальнейшем будет использовано для оценки эффективности работы нефтяных скважин на основе собственных характеристик нефтяных скважин и влияющих показателей работы нагнетательных скважин.

### Литература

1. Вирстюк А. Ю. Разработка модуля по контролю нагнетательных скважин с неравномерным профилем приемистости // Север России: стратегии и перспективы развития : материалы II Всерос. науч.-практич. конф. 2016. Сургут, 2016. С. 140.
2. Вирстюк А. Ю., Микшина В. С. Применение математических моделей для оценки эффективности работы добывающих скважин // Информационная безопасность регионов России (ИБРР-2017) : материалы юбилейной X Санкт-Петербургской межрегион. конф. СПб., 1–3 ноября 2017 г. СПб. : СПОИСУ, 2017. С. 373–374.
3. Тихомиров Н. П. Методы эконометрики и многомерного статистического анализа. М. : Экономика, 2011. 647 с.
4. Токарев М. А., Чинаров А. С. Статистические методы прогноза нефтеотдачи и оценки эффективности воздействия на пласт : моногр. Уфа, 2007. 96 с.
5. Халафян А. А. Statistica 6: стат. анализ данных. М. : БиномПресс, 2007. 512 с.
6. Basilevsky A. Statistical factor analysis and related methods: theory and applications. John Wiley & Sons, Inc., 2008. 737 p.
7. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second Edition. Springer, 2009. 764 p.
8. James G., Hastie T., Tibshirani R., Witten D. An Introduction to Statistical Learning: With Applications in R. Springer, 2013. 426 p.
9. Marques de Sá J. P.. Applied Statistics Using SPSS, Statistica, MATLAB and R. Springer, 2007. 505 p.
10. Kettaneh N., Berglund. A. PCA and PLS with very large data sets // Computational Statistics & Data Analysis. 2005. Vol. 48. P. 69–85.