

# ПРОБЛЕМЫ МЕТОДОЛОГИИ

УДК 14.01.77

А. К. Колесников,  
И. П. Лебедева

## ПРОБЛЕМА ИНТЕРПРЕТАЦИИ РЕГРЕССИОННЫХ МОДЕЛЕЙ В ПЕДАГОГИЧЕСКОМ ИССЛЕДОВАНИИ

*Аннотация.* В статье рассматриваются возможности использования регрессионных моделей для анализа результатов педагогического эксперимента, дается оценка их адекватности и научно-педагогическая интерпретация.

*Ключевые слова:* математические модели, педагогический эксперимент, регрессионный анализ, научное исследование, педагогическая интерпретация

*Abstract.* The paper considers the possible usage of regression patterns when analysing the results of pedagogical experiments, evaluation of their identity done and scientific pedagogical interpretation given.

*Index terms:* mathematical patterns, pedagogical experiments, regression analysis, scientific research, pedagogical interpretation.

Математическое моделирование в научном исследовании используется на разных этапах его реализации, выполняя важную роль в интеграции эмпирического и теоретического уровней познания, необходимой для выявления причинно-следственных отношений [3, 4]. В педагогическом исследовании особое значение имеет обоснованность выводов о причинных связях, которые устанавливаются в процессе эксперимента (т. е. определяется, вызывает ли изменение одной переменной изменение другой переменной). Эксперимент позволяет манипулировать одной переменной и в условиях тщательного контроля наблюдать ее влияние на другую интересующую исследователя переменную. Это влияние описывается в виде некоей функциональной зависимости.

Постановка эксперимента является довольно сложным процессом, поскольку необходимо учитывать множество факторов, способных повлиять на точность и научную ценность результата. В педагогическом исследовании учесть все факторы невозможно, важно среди них правильно выделить существенные. Разработано множество способов исключения влияния посторонних переменных в условиях эксперимента. Однако невозможно добиться абсолютного контроля исследуемого фактора и полностью нивелировать воздействие случайных факторов. В итоге получен-

ный результат оказывается приближительным, а его точность требует специального оценивания.

Аналитической формой представления причинно-следственных отношений являются регрессионные модели. Научная обоснованность и популярность регрессионного анализа делает его одним из основных математических средств моделирования явлений и процессов. В педагогических исследованиях этот метод, как правило, применяется для сглаживания экспериментальных данных и получения количественных оценок сравнительного влияния различных факторов на результативную переменную. Учитывая существование большого количества этих факторов, мы считаем наиболее интересным применение многофакторной регрессии.

Однако потенциальные возможности регрессионных моделей остаются нереализованными, поскольку достаточно сложно выполнить все требования, предъявляемые к данным. Основными проблемами множественной регрессии являются размерность и мультиколлинеарность. Исследователь постоянно вынужден искать баланс между простотой модели и полнотой описания исследуемого явления, одновременно удовлетворяя необходимым математическим критериям адекватности модели.

В классическом варианте регрессионного анализа ставится задача поиска аналитической зависимости для полученных в эксперименте данных. Выделяют следующие этапы моделирования:

- предварительный анализ данных;
- формирование набора моделей (определение вида функций), подходящих для описания исследуемого явления;
- численное оценивание параметров;
- определение адекватности моделей;
- выбор лучшей модели;
- получение точечного и интервального прогнозов;
- верификация прогноза (сопоставление расчетных результатов по модели с соответствующими данными действительности) [3].

В процессе реализации указанных этапов необходимо учитывать, во-первых, что при построении модели делается ряд статистических предположений. Их истинность часто остается под вопросом. Поэтому уместным вопросом будет не «идеально ли модель согласуется с данными», а «согласуется ли она достаточно хорошо, чтобы получить новую информацию об объекте исследования, объяснить структуру данных и быть полезной для практического применения». Во-вторых, идеальное соответствие модели данным не означает, что она верна. Если модель верна, то она согласуется с данными. Обратное неверно. Если модель согласуется с данными, то она не обязательно верна. Можно построить другие модели, так

же хорошо согласуемые с исходными данными. Возникает проблема правильной интерпретации, определения научной и практической ценности полученной модели и возможностей прогнозирования на ее основе.

Одно из основных назначений регрессионных моделей – прогнозирование изменений зависимой переменной при соответствующих вариациях независимой переменной. Однако их прогнозные качества весьма ограничены, поскольку данные не имеют временной структуры (измеряются однократно). Поэтому применение регрессионных моделей с целью прогнозирования целесообразно лишь по отношению к универсальным обобщающим и статистическим показателям, взаимосвязи между которыми заведомо имеют устойчивый характер. Но даже в этом случае прогноз получается грубый и приближенный. Эффективным способом повышения прогнозных свойств регрессионной модели является двукратное измерение одного и того же показателя. Тогда в качестве переменных рассматриваются не статические показатели, а динамические (отклонения). При этом возникают ситуации, когда целесообразно использовать в уравнении одновременно и статические, и динамические показатели, с тем чтобы учитывать как достигнутый уровень (экстраполировать статику), так и намечающиеся тенденции (экстраполировать динамику). Чаще всего в качестве динамического показателя выбирают зависимую переменную, например, отклонение эффективности учебной деятельности. Тогда для факторов, влияющих на эту эффективность (способности, мотивы, посещаемость и др.), будет оцениваться не вклад в изменение независимой переменной, а то, в каком направлении будет меняться ее отклонение: в сторону увеличения или уменьшения.

После получения адекватной модели с точки зрения математических критериев оценивают ее так называемую «содержательную адекватность», или целесообразность. Однако даже если модель адекватна в содержательном плане, остается проблема ее интерпретации. Какой смысл имеют полученные соотношения? Что дают полученные оценки? Каковы рекомендации по их использованию?

Например, построены регрессионные уравнения, в которых рассматривалось влияние количества пропусков студентами занятий на успеваемость. Получены интересные количественные оценки. Если студент пропустит больше половины занятий, то ему почти гарантированы (с достаточной вероятностью) 2,5 балла на экзамене (по пятибалльной шкале). Если он не пропустит ни одного занятия, ему почти гарантированы четыре балла. Этот результат не имеет научной ценности и может рассматриваться как практически полезная модель, иллюстрирующая корректность применения регрессионного анализа. Очевидно, что есть другие существ-

венные факторы, влияющие на успеваемость. Какой из них доминирует, в каком они находятся соотношении для данной категории испытуемых и конкретной учебной дисциплины? Каково их оптимальное соотношение? Ответы на эти вопросы могут представлять определенный интерес. Тогда наиболее перспективно использование множественной регрессии.

Следуя принципу простой структуры и опираясь на идею линеаризации функции, естественно начинать анализ явления с построения наиболее просто интерпретируемых линейных моделей. Рассматривая их совокупность, можно получить значительно больше информации, чем на основе одного уравнения.

Например, получены регрессионные уравнения ( $R > 0,8$ ;  $p < 0,01$ ):

$$\mathcal{E}_w = 37,2 + 1,3A_1 + 0,6A_2 - 0,1A_3;$$

$$\mathcal{E}_k = 46,3 - 0,4A_1 + 2,6A_2 + 0,2A_3;$$

$$\mathcal{E}_m = 34,8 - 0,2A_1 + 1,1A_2 + 3,4A_3,$$

где  $\mathcal{E}_w$ ,  $\mathcal{E}_k$ ,  $\mathcal{E}_m$  – эффективность усвоения текста по результатам решения задач трех групп сложности (индексом  $w$  обозначена группа задач, требующих действий по алгоритму;

$k$  – комбинирования известных способов;

$m$  – проявления креативности);

$A_1$ ,  $A_2$ ,  $A_3$  – активность интеллекта, соответствующая первому, второму и третьему уровням.

Коэффициенты уравнения позволяют сравнить долю разного вида активности для усвоения текста на разном уровне (воспроизведения материала, комбинирования его элементов или творческого применения в незнакомых ситуациях). Таким образом, глубина изучения текста в значительной мере определяется соотношением уровней интеллектуальной активности при работе с ним: с увеличением сложности задач не только меняется доминирующий уровень активности, но и существенно увеличивается ее «вес» [5].

Интерпретация линейной модели проводится с учетом знаков коэффициентов регрессии, их величины и соотношений между ними. Причем последние интересны в случае сопоставления нескольких уравнений. Если в качестве зависимой переменной использовать отклонения, например, эффективности усвоения текста в приведенном выше примере, тогда можно было бы сделать вывод о влиянии разных уровней активности не на успешность решения задач различной сложности, а на изменение этой успешности. В результате не только повышаются прогнозные свойства модели, но и расширяется смысл выявленных взаимосвязей.

В рассмотренном примере дана педагогическая интерпретация зафиксированным в регрессионных уравнениях количественным соотношениям. Она допустима только после того как проведена оценка адекватности модели с точки зрения математических критериев в терминах статистических гипотез. Для линейной модели – это критерий Фишера (для установления адекватности модели в целом) и критерий Стьюдента (для установления достоверных отличий каждого коэффициента регрессии от нуля). Кроме того, требуется обоснование построения именно данного вида уравнения: отсутствие значимых корреляций между независимыми переменными, наличие значимых корреляций между каждой независимой переменной в отдельности и результативным признаком, полнота описания исследуемого явления данной совокупностью переменных и их однородность.

Предложенная педагогическая (научно-теоретическая) интерпретация результатов моделирования выполнена с позиции авторской дидактической теории взаимодействия ученика и предмета изучения. Наличие научной новизны в полученных соотношениях позволяет рассматривать их с точки зрения практической значимости и дать соответствующие рекомендации для образовательной практики. Это определяет следующий уровень интерпретации – научно-популярный [5].

В данном случае предмет исследования допускает линеаризацию: содержательная интерпретация является целесообразной. Однако границы применимости линейных моделей значительно уже, чем нелинейных. Полученные соотношения связаны с конкретной ситуацией (условиями проведения эксперимента), т. е. модель имеет локальный характер. Как добиться ее большей универсальности? Возможны два пути: воспроизведение эксперимента в других условиях и получение нелинейной модели. В первом случае можно добиться большей обоснованности линейной структуры взаимосвязи, ее устойчивости и расширить границы применимости модели. При этом дополнительная информация не будет получена. Второй путь более сложный. Нелинейная модель только с тремя независимыми переменными достаточно громоздка. Это существенно затрудняет оперирование регрессионным уравнением и его интерпретацию. Для проведения последней необходимо провести исследование функции, привлекая серьезный математический аппарат. И если удалось получить новую информацию (экстремум функции, наибольшее или наименьшее значения и т. д.), то возникает потребность в ее содержательной интерпретации. Естественно задать вопрос: могут ли адекватные, построенные для одних и тех же начальных условий модели отражать разные тенденции, давать противоречащие друг другу интерпретации?

При выполнении всех предпосылок регрессионного анализа такого не может быть. Если учесть диапазон изменения независимых переменных и зависимой переменной, то возможные интерпретации будут либо дублировать, либо дополнять друг друга. Проблема может оказаться другая: отсутствие адекватных моделей с точки зрения математических критериев или их содержательного смысла. Можно сделать вывод, что регрессионные модели не подходят для описания исследуемого явления. Тогда возникает вопрос: эксперимент неправильно поставлен, его результат отрицательный, т. е. не существует функции, описывающей влияние исследуемого фактора или факторов на результативный признак?

В этом случае полезно возвратиться к первому этапу моделирования – анализу данных. Проблема сводится к структурированию данных: как переменных, так и случаев (наблюдений). Речь идет о формализации предмета исследования, в основе которой его представление с помощью системы признаков, удовлетворяющей таким условиям, как полнота, экономность, структурированность, количественная определенность. Выполнение второго и третьего условий предполагает использование специальных математических методов, обеспечивающих необходимую группировку или интеграцию данных (например, факторного и кластерного анализа).

В случае же получения адекватных, достаточно точных и хорошо интерпретируемых моделей их параметры можно рассматривать как исходные данные для продолжения структурно-количественного анализа исследуемого объекта. Поэтому в современных исследованиях регрессионный анализ рассматривается не изолированно, а в комплексе с другими математическими методами. Обозначим некоторые направления его комплексного применения:

1. Использование в качестве зависимой переменной или независимых переменных факторов, выделенных в результате факторного анализа. Причем смысл этих факторов может быть разный: интегрированные или латентные переменные. Это в значительной мере определяет содержательную интерпретацию модели.

2. Построение систем линейных уравнений, отражающих многоаспектность исследуемого объекта.

3. Использование коэффициентов регрессионного уравнения или их соотношений в качестве параметров при дальнейшей обработке результатов эксперимента.

4. Сопоставление регрессионных уравнений одного и того же вида в разных выборках или для разных условий. При этом сравниваются не сами коэффициенты регрессии, а их соотношения.

Таким образом, становится актуальной разработка специальных схем анализа результатов эксперимента, учитывающих объективную существующую иерархию объекта исследования.

Простейшие варианты регрессионного анализа в последнее время достаточно широко применялись в педагогических исследованиях [1, 2, 6]. Полученные метрические соотношения использовались для разработки различных технологий и методик обучения. На современном этапе развития математических методов в значительной мере изменилась постановка исследовательских задач. Они связаны не только с поиском аналитической зависимости экспериментальных данных, но и с возможностью изучения причинных связей на основе анализа конкретной регрессионной модели, в которой значения коэффициентов изначально заданы. Ряд авторов считает, что такие статистические процедуры эффективны при использовании неэкспериментальных данных (например, при моделировании поведения) и являются альтернативой оперирования с экспериментальными данными.

Свое дальнейшее развитие регрессионный анализ получил в связи с применением техники многомерного анализа – моделирования структурных уравнений. Одной из основных идей, положенных в основу такого моделирования, является изучение эффекта аддитивных и мультипликативных преобразований. Смысл заключается в том, что можно проверить, связаны ли переменные линейной зависимостью, изучая их дисперсии и ковариации. По мнению Р. Корсини, А. Ауэрбаха, предметом многомерного исследования являются гипотетические конструкты, не доступные прямому наблюдению, но оказывающие решающее воздействие на измеряемые переменные и характер связи между ними. Латентный аспект таких переменных учитывают структурные уравнения, которые позволяют не только получить информацию о степени эмпирической связи между переменными, но и провести исследование причинных связей при данном эмпирическом объединении. При выполнении определенных условий уравнения регрессии используются для оценки структурных уравнений.

Проблема их интерпретации становится еще более актуальной, поскольку рассматривается система уравнений для конкретной ковариационной структуры. А экспериментальные данные анализируются с точки зрения соответствия ей. Если это соответствие установлено правильно, то систему уравнений принимают в качестве предварительной гипотезы о причинно-следственных отношениях.

Таким образом, математический аппарат, связанный с использованием регрессионных моделей, становится все более сложным и совершенным. Но его возможности реализуются незначительно, поскольку до сих

пор остается серьезной научной проблемой разумное сочетание содержательного и формального подходов в педагогическом исследовании [4]. Решение этой проблемы позволит создать необходимую математико-педагогическую основу для получения адекватных моделей, обладающих научной ценностью и практической значимостью.

### Литература

1. Александров Г. Н., Белогуров А. Ю. Математические методы в психологии и педагогике. Владикавказ: Изд-во Сев.-Осет. гос. ун-та, 1997. 302 с.
2. Бартоломью Д. Стохастические модели социальных процессов: пер. с англ. / под ред. О. В. Старовойта. М.: Финансы и статистика, 1985. 295 с.
3. Введение в математическое моделирование: учеб. пособие / под ред. П. В. Трусова. М., 2005.
4. Загвязинский В. И., Атаханов Р. Методология и методы психолого-педагогического исследования. М.: Академия, 2001. 208 с.
5. Лебедева И. П. Теория взаимодействия систем «ученик» и «объект изучения». Пермь: Изд-во Пермск. ун-та, 2001.
6. Мизинцев В. П. Применение моделей и методов моделирования в дидактике. М.: Знание, 1977. 50 с.