



В. А. Дорофеев
Ю. А. Мочалова

ОСНОВЫ РЕГРЕССИОННОГО МОДЕЛИРОВАНИЯ ДЛЯ ПСИХОЛОГОВ

Учебное пособие
по дисциплине
«Математическая статистика
и математические методы
в психологии»



Вадим Дорофеев

**Основы регрессионного
моделирования для психологов**

«Южный Федеральный Университет»

Дорофеев В. А.

Основы регрессионного моделирования для психологов /
В. А. Дорофеев — «Южный Федеральный Университет»,

ISBN 978-5-92-752549-2

Содержательно учебное пособие дополняет лекционно-практический курс «Математическая статистика и математические методы в психологии» и отражает попытку авторов представить учебный материал по регрессионному моделированию в психологии на основе синтеза методологических аспектов психологии, основ статистико-математических технологий регрессионного анализа и компьютерных технологий обработки эмпирических результатов. Предназначено для студентов и магистрантов психологических и психолого-педагогических специальностей.

ISBN 978-5-92-752549-2

© Дорофеев В. А.
© Южный Федеральный Университет

Содержание

Введение	5
Тема 1. Методологические аспекты регрессионного анализа в психологии	7
1.1. Законы и закономерности в психологии и регрессионный анализ	7
Конец ознакомительного фрагмента.	12

В. А. Дорофеев, Ю. А. Мочалова

Основы регрессионного моделирования для психологов

Введение

Разработка данного учебного пособия является следствием ряда формальных и содержательных проблем, которые имеют место в ходе изучения студентами-психологами такого раздела учебной дисциплины «Математическая статистика и математические методы в психологии», как «Регрессионное моделирование в психологии».

Во-первых, разработка компьютерных статистических пакетов для обработки эмпирических результатов в гуманитарных дисциплинах (SPSS, Statistica и др.) не только привела к целому ряду позитивных моментов в научно-исследовательской деятельности в психологии (например, значительно сократилось время статистической обработки эмпирических данных, появилась возможность вносить определенные коррективы в планирование самого эмпирического исследования с опорой на промежуточные статистические результаты и т. д.), но и породила некоторые негативные моменты. В частности, можно привести два таких момента. Первый связан с тем, что простота «нажатия кнопок по схеме» в компьютерных статистических пакетах, приводящего к конечному продукту статистического анализа (в Интернете в режиме свободного доступа достаточно много информационных источников, презентующих эту схему), стала проявляться в том, что в научно-психологических исследованиях статистический анализ стали использовать «где надо» и «где не надо», забывая о том, что психология имеет весьма специфичный объект исследования. К примеру, многие в своих исследованиях пытаются найти некие статистические обобщения с применением компьютерных статистических пакетов (благо доступ к таким программам стал «беспроблемным») даже там, где с методологической точки зрения речь идет об уникальности бытия человеческой личности (понимающая методология). Особенно эта тенденция проявляется в исследовательских работах студентов-психологов, ориентированных на формальные требования к научно-исследовательским работам (требование использовать статистические методы). Второй связан с тем, что психолога-исследователя стали интересовать только «конечные распечатки» и он перестал обращать внимание на технологические аспекты математико-статистических расчетов, осуществляемых компьютером. Как следствие, абсолютно перестал учитываться тот факт, что в технологиях математико-статистического анализа для их проведения существует достаточно много формальных условий-требований, автоматически проявляющихся в «конечных распечатках», но выполнение которых не всегда возможно применительно к конкретной научной дисциплине, в частности к психологии.

Во-вторых, педагогический опыт показывает, что когда речь идет об изучении студентами-психологами продуктов моделирования в психологии по разным учебным дисциплинам, то приоритет отдается моделям, которые основаны на концептуальных взглядах автора модели, а не моделям, основанным на строгом статистико-математическом (мы специально используем связку в виде *статистико-математический*, а не *математико-статистический*, потому что в регрессионном моделировании математические расчеты осуществляются на основе статистических данных) анализе реальных эмпирических результатов (следствие – *низкая объективность (научность)* продуктов моделирования в нынешней психологии). Одна из причин такого положения вещей кроется в отсутствии у студентов понимания сути и технологических составляющих проведения регрессионного моделирования (основанного на строгом

статистико-математическом анализе реальных эмпирических результатов) в психологических исследованиях.

Учитывая вышесказанное, в данном пособии мы попытались доступным для понимания языком представить содержание, которое синтезировало бы содержательные аспекты трех составляющих (компонентов):

1) методологических основ психологии как науки; 2) математических основ регрессионного анализа; 3) компьютерных технологий математико-статистического анализа эмпирических результатов.

Следуя этим путем, мы понимали, что создание такого «гибрида», как и любого гибрида, обязательно приведет и к негативным последствиям. В частности, он утратит часть содержательных аспектов каждого из своих «родительских» компонентов по сравнению с ситуацией, если бы мы рассматривали их в отдельности (чтобы снизить влияние этого фактора, в пособии в сносках приводится литература, которая позволит повысить знания в каждом из этих компонентов). Несмотря на это, как нам представляется, изучение содержания данного пособия студентами-психологами позволит решить две важные учебные задачи. Во-первых, будет способствовать развитию у них стратегического видения возможности применения регрессионного моделирования в реализации своих научно-исследовательских проектов (численное моделирование дает наиболее точную прогнозируемую оценку изучаемым явлениям). Во-вторых, будет способствовать развитию у них способности более качественно проводить регрессионное моделирование с технологической точки зрения.

Автором предисловия и параграфов 1.1, 1.3, 2.2, 2.3, 2.4, 3.1, 3.2, 3.3, 3.4, 4.1, 4.3 и тестовых заданий для самопроверки является В. А. Дорофеев, автором параграфов 1.2, 2.1, 4.2 – Ю. А. Мочалова.

Тема 1. Методологические аспекты регрессионного анализа в психологии

1.1. Законы и закономерности в психологии и регрессионный анализ

Чтобы иметь статус науки, психология должна опираться на определенные законы и закономерности. В задачи данного пособия не входит рассматривать методологические аспекты наличия или отсутствия законов и закономерностей в психологии и соотношения между ними¹, только заметим, что в основе закона лежит связь между одним или несколькими явлениями (переменными) и другим или другими явлениями (переменными). Если следовать всем канонам требований к научности (объяснительности) результатов исследования, то разговор идет прежде всего о связи каузальной.

Что касается определения сущности каузальной связи, то стоит заметить, что единого подхода к пониманию этой сущности не существует и, как следствие этого, нет единого подхода к ее эмпирическому изучению².

С содержательной стороны построения регрессионных моделей, освещаемой данным пособием, важно заметить, что в определение каузальной связи может включаться или не включаться такой фактор, как время. Если фактор времени включить в определение, тогда каузальную связь можно определить следующим образом.

Каузальная связь (от лат. *causa* – причина) – одна из важнейших форм взаимосвязи и взаимообусловленности явлений и процессов бытия, выражающая такую генетическую связь между ними, при которой одно явление (процесс), называемое причиной, при наличии определенных условий неизбежно порождает, вызывает к жизни другое явление (процесс), называемое следствием (или действием).

При каузальной связи причина и следствие разнесены во времени (одна психологическая переменная (причина) появляется раньше другой (следствие), а не наоборот). Следствие не может быть раньше причины – причина и следствие связаны генетически.

Например, в психологии в конце 1930-х гг. была сформулирована ставшая впоследствии широко известной в психологической науке теория фрустрации – агрессии. Ее авторами являются Н. Миллер, Д. Доллард, М. Дуб, Д. Маурер и Р. Сизерс.

Представители этого направления Н. Миллер и Д. Доллард следующим образом сформулировали гипотезу: наличие агрессивного поведения всегда предполагает существование фрустрации и, наоборот, существование фрустрации всегда ведет к некоторой форме агрессии. Таким образом, в данной модели ее авторы отразили каузальную связь, включающую фактор времени: сначала фрустрация → потом агрессия.

Однако экспериментальные данные в рамках вышеназванной теории показали, что только одной теории недостаточно для объяснения агрессивного поведения. Реальные экспериментальные исследования показали, что существуют явные и неявные переменные, которые не только искажают генетическую связь между фрустрацией и агрессией, но и нарушают универсальность каузальной (генетической) направленности такой связи.

¹ См. например: Корнилова Т. В., Смирнов С. Д. Методологические основы психологии: учеб. пособие. М.; СПб., Питер, 2006. С. 92–113.

² См. например: Митина О. В. Методы исследования каузальных связей. URL: http://psyjournals.ru/exp_collection/issue/33028_full.shtml.

Применительно к задачам и содержанию данного пособия заметим, что теория вероятностей и математическая статистика, которые в нем представлены, – лишь инструмент для изучения статистической зависимости между явлениями, но *однозначно* не позволяют установить наличие каузальной связи, включающей фактор времени. В психологии представления о генетически определяемой каузальной связи должны быть привнесены из некоторой другой теории, которая позволяет содержательно объяснить изучаемое явление.

Заметим, что психология как научная дисциплина в отношении включения фактора времени в регрессионные модели находится в весьма затруднительном положении по сравнению с другими дисциплинами (например, экономикой, биологией). В экономике и биологии в регрессионной модели фактор времени включается в моделирование в виде статистических данных, собранных на протяжении определенных предшествующих интервалов времени, чего практически не встретишь в психологии. Причин тому несколько. Например, одной из причин такого положения вещей (невозможность сформировать статистические базы данных на определенных периодах развития изучаемых явлений) является отсутствие в психологии надежного измерительного инструментария (за исключением психологии ощущения и восприятия (сантиметры, секунды)), позволяющего хотя бы отчасти сохранить объективность результатов измерений.

Иногда в литературе можно обнаружить исследования, в которых психологи-исследователи для нахождения каузальной связи используют дисперсионный анализ (ANOVA, MANOVA). Но дисперсионный анализ позволяет определить *статистическую достоверность влияния* одной (нескольких) переменной на зависимую (зависимые) переменную (проявляется в мере дисперсии зависимой переменной), однако наличие статистически достоверного влияния является необходимым, но еще далеко не достаточным условием генетически обусловленной связи между явлениями.

Таким образом, выделим *первый важный аспект регрессионного анализа*: он не позволяет устанавливать связь каузальную, понимаемую как связь генетическая между явлениями и процессами. Например, установить факт того, что наличие состояния фрустрации всегда приведет к агрессии.

Тогда возникает вопрос: какую же связь позволяет находить регрессионный анализ?

В ответе на этот вопрос дадим характеристику той связи, с которой имеет дело психолог-исследователь после проведения эмпирического исследования, когда выполнены все требования к технологиям сбора эмпирических результатов и соблюдены требования к объему статистической выборки.

Если выполнены вышеназванные условия (соблюдены требования к технологиям сбора эмпирических результатов и к объему статистической выборки) и полученные эмпирические результаты нанесены на двумерный график, то мы столкнемся с тем, что всегда одним и тем же значениям одной переменной будут соответствовать разные значения другой переменной.

На рис. 1.1 представлены два возможных варианта графического представления такой ситуации.

Связь, которая представлена на рис. 1.1, называется вероятностной (стохастической).

Стохастическая связь – связь, при которой каждому значению одной переменной значение других переменных соответствует не однозначно, а с определенной долей вероятности.



Рис. 1.1. Варианты диаграмм совместного рассеивания точек (каждая точка – испытуемый) в двумерном исследовании

Количественным выражением такого вида связи является коэффициент корреляции.

При стохастической связи переменные как случайные величины заданы совместным распределением вероятностей величины.

Не вдаваясь в объяснение статистических технологий решения задачи о нахождении количественного выражения данного вида связи (коэффициента корреляции)³, охарактеризуем основной *недостаток стохастической связи* для объяснительного (научного) подхода в исследовательской деятельности.

Он заключается в том, что у нас нет никакой возможности даже с определенной долей вероятности спрогнозировать конкретную количественную выраженность одной переменной при условии, что вторая переменная будет также принимать конкретные количественные значения.

Для того чтобы решить эту задачу, необходимо перейти к другому виду выражения этой связи – математическому, позволяющему отражать эту связь в виде определенной математической функции (функциональная связь).

Функциональная связь – связь, при которой каждому значению одной переменной соответствует одно и только одно значение другой переменной, и эта связь выражается какой-либо математической функцией.

Таким образом, выделим *второй важный аспект регрессионного анализа*: он позволяет устанавливать связь между переменными, выражаемую определенной математической функцией, что дает возможность спрогнозировать конкретную количественную выраженность одной переменной при условии, что вторая переменная будет также принимать конкретные количественные значения.

Мы обращаем внимание на тот факт, что когда разговор идет о функциональной связи, подразумевается, что она выражается определенной математической функцией. Различные виды математических функций, которые используют в регрессионном анализе, будут приведены ниже.

В таких рамках регрессионный анализ направлен на поиск ответов на два макровопроса.

Первый – определение возможного наличия связи между переменными, которую можно выразить математической функцией на базе вероятностной (стохастической) связи и которая всегда имеет место в психологии при достаточном объеме эмпирической выборки.

Второй: если эта связь существует, то какой вид математической функции ее отражает?

Поясним на примере. В психологии давно общепризнана закономерность, которая получила название закона Йеркса–Додсона и которая гласит: «По мере увеличения интенсивности мотивации качество деятельности изменяется по колоколообразной кривой: сначала повышается, затем, после перехода через точку наиболее высоких показателей успешности, постепенно

³ В прил. 1 представлена краткая схема расчетов и анализа коэффициента линейной корреляции Пирсона r_{xy} .

снижается». Графически, в самом общем виде, это выглядит как перевернутая парабола (рис. 1.2).



Рис. 1.2. Кривая, отражающая закон Йеркса–Додсона

Как показывает опыт преподавания в вузе, подавляющее число студентов, когда видят эту кривую, не совсем понимают, что представлен вариант кривой, на которой отражена лишь *усредненная* тенденция изменения значений показателя качества деятельности (зависимая переменная) при изменении значений интенсивности мотивации (независимая переменная).

Можно предполагать, что когда реальные экспериментальные данные наносились на график, в котором по оси абсцисс откладывалось значение такого параметра, как уровень мотивации (в эксперименте – независимая переменная), а по оси ординат – значение такого параметра, как качество деятельности (в эксперименте – зависимая переменная), то график в действительности имел *приблизительно* следующий вид (на графике отражена тенденция – одному и тому же значению интенсивности мотивации соответствует несколько значений качества деятельности – это получается в силу того, что на один и тот же стимул разные испытуемые показывали разные результаты) (рис. 1.3).

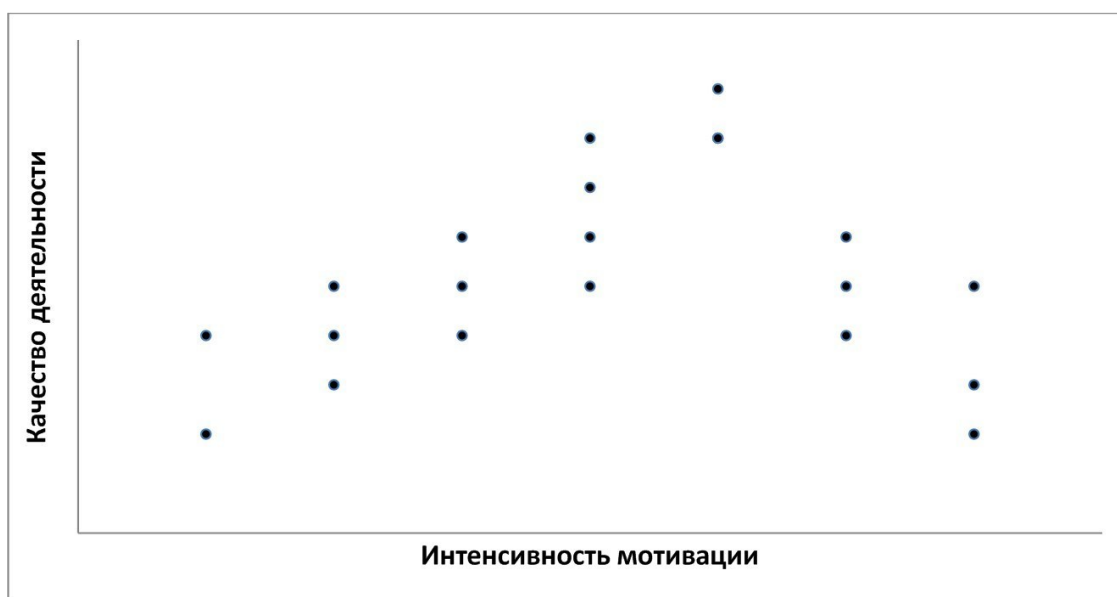


Рис. 1.3. Модель варианта реальных экспериментальных данных в законе Йеркса–Додсона

В реальности связь между двумя переменными носила не функциональный характер, который отражен квадратичной функцией (параболой), а стохастический, выраженный графиком, по форме напоминающим параболу. И только в результате аппроксимации регрессией реальных данных была получена параболическая функциональная зависимость, показывающая, как изменяется *в среднем*

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.