***9.13. Корреляционно-регрессионные модели и их применение в анализе и прогнозе***

Корреляционно-регрессионной моделью (КРМ) системы взаимосвязанных признаков является такое уравнение регрессии, которое включает основные факторы, влияющие на вариацию результативного признака, обладает высоким (не ниже 0,5) коэффициентом детерминации и коэффициентами регрессии, интерпретируемыми в соответствии с теоретическим знанием о природе связей в изучаемой системе. Приведенное определение КРМ включает достаточно строгие условия: далеко не всякое уравнение регрессии можно считать моделью. В частности, полученное выше по 16 хозяйствам уравнение не отвечает последнему требованию из-за противоречащего экономике сельского хозяйства знака при

факторе х2 — доля пашни. Однако в учебных целях будем рассматривать его как модель. Теория и практика выработали ряд рекомендаций для построения корреляционно-регрессионной модели.

1. Признаки-факторы должны находиться в причинной связи с результативным признаком (следствием). Поэтому недопустимо, например, в модель себестоимости у вводить в качестве одного

из факторов xj коэффициент рентабельности, хотя включение такого «фактора» значительно повысит коэффициент детерминации.

2. Признаки-факторы не должны быть составными частями результативного признака или его функциями.

3. Признаки-факторы не должны дублировать друг друга, т.е. быть коллинеарными (с коэффициентом корреляции более 0,8). Так, не следует в модель производительности труда включать энерго- и фондовооруженность рабочих, поскольку эти факторы тесно связаны друг с другом в большинстве объектов.

4. Не следует включать в модель факторы разных уровней иерархии, т.е. фактор ближайшего порядка и его субфакторы. Например, в модель себестоимости зерна не следует включать и урожайность зерновых культур, и дозу удобрений под них или затраты на обработку гектара, показатели качества семян,

плодородия почвы, т.е. субфакторы самой урожайности.

5. Желательно, чтобы для результативного признака и факторов соблюдалось единство единицы совокупности, к которой они отнесены. Например, если у — валовой доход предприятия, то и все факторы должны относиться к предприятию: стоимость производственных фондов, уровень

специализации, численность работников и т.д. Если же у — средняя зарплата рабочего на предприятии, то факторы должны относиться к рабочему: разряд или классность, стаж работы, возраст, уровень образования, энерговооруженность и т.д. Правило это некатегорическое, в модель заработной платы

рабочего можно включить, к примеру, и уровень специализации предприятия. Вместе с тем нельзя забывать о предыдущей рекомендации.

6. Математическая форма уравнения регрессии должна соответствовать логике связи факторов с результатом в реальном объекте. Например, такие факторы урожайности, как дозы разных удобрений, уровень плодородия, число прополок и т.п., создают прибавки величины урожайности, малозавися-Аше

друг от друга; урожайность может существовать и без любого из этих факторов. Такому характеру связей отвечает аддитивное уравнение регрессии:



Первое слагаемое в правой части равенства — это отклонение, которое возникает за счет отличия индивидуальных значений факторов у данной единицы совокупности от их средних значений по совокупности. Его можно назвать эффектом факторообеспеченности. Второе слагаемое — отклонение, которое возникает за счет не входящих в модель факторов и отличия индивидуальной эффективности факторов у данной

единицы совокупности от средней эффективности факторов в совокупности, измеряемой коэффициентами условно-чистой регрессии. Его можно назвать эффектом

фактороотдачи.

Таблица 9.12 Анализ факторообеспеченности и фактороотдачи

по регрессионной модели уровня валового дохода



**Пример.** Рассмотрим расчет и анализ отклонений по ранее построенной модели уровня валового дохода в 16 хозяйствах. Знаки тех и других отклонений 8 раз совпадают и 8 раз не совпадают. Коэффициент корреляции рангов отклонений двух видов составил 0,156. Это означает, что связь вариации факторообеспеченности с вариацией фактороотдачи слабая, несущественная (табл. 9.12). Обратим внимание на хозяйство № 15 с высокой факторообеспеченностью (15-е место) и самой худшей фактороотдачей (1-й ранг), из-за которой хозяйство недополучило по 1 22 руб. дохода с 1 га. Напротив, хозяйство № 5 имеет факторообеспеченность ниже средней, но благодаря более эффективному использованию факторов получило на 125 руб.

дохода с 1 га больше, чем было бы получено при средней по совокупности эффективности факторов. Более высокая

эффективность фактора х\ (затраты труда) может означать

более высокую квалификацию работников и большую

заинтересованность в качестве выполняемой работы. Более высокая эффективность фактора хз с точки зрения доходности может заключаться в высоком качестве молока (жирность, охлажденность), благодаря которому оно реализовано по более высоким ценам. Коэффициент регрессии при х2, как уже отмечено, экономически не обоснован. Использование регрессионной модели для прогнозирования состоит в подстановке в уравнение регрессии ожидаемых значений факторных признаков для расчета точечного прогноза результативного признака или (и) его доверительного интервала с заданной вероятностью, как уже сказано в 9.6. Сформулированные там же ограничения прогнозирования по уравнению регрессии сохраняют свое значение и для многофакторных моделей. Кроме того, необходимо соблюдать системность между подставляемыми в модель значениями факторных признаков.

Формулы расчета средних ошибок оценки положения гиперплоскости регрессии в заданной многомерной точке и для индивидуальной величины результативного признака весьма сложны, требуют применения матричной алгебры и здесь не рассматриваются. Средняя ошибка оценки значения результативного признака, рассчитанная по программе ПЭВМ «Mi-crostat» и приведенная в табл. 9.7, равна 79,2 руб. на 1 га. Это лишь среднее квадратическое отклонение фактических значений дохода от расчетных по уравнению, не учитывающее ошибки положения самой гиперплоскости регрессии при экстраполяции значений факторных признаков. Поэтому ограничимся точечными прогнозами в нескольких вариантах (табл. 9.13).

Для сравнения прогнозов с базисным уровнем средних по совокупности значений признаков введена первая строка таблицы. Краткосрочный прогноз рассчитан на малые изменения факторов за короткое время и снижение трудообеспеченности.

Таблица 9.13 Прогнозы валового дохода по

регрессионной модели



Результат неблагоприятен: доход снижается. Долгосрочный прогноз А — «осторожный», он предполагает весьма умеренный прогресс факторов и соответственно небольшое увеличение дохода. Вариант Б — «оптимистический», рассчитан на существенное изменение факторов. Вариант 5 построен по способу, которым Агафья Тихоновна в комедии Н. В. Гоголя

«Женитьба» мысленно конструирует портрет «идеального

жениха»: нос взять от одного претендента, подбородок от

другого, рост от третьего, характер от четвертого; вот если бы соединить все нравящиеся ей качества в одном человеке, она бы не колеблясь вышла замуж. Так и при прогнозировании мы объединяем лучшие (с точки зрения модели дохода) наблюдаемые значения факторов: берем значение Х[ от хозяйства № 10, значение х2 от хозяйства № 2, значение х3 от хозяйства № 16. Все эти значения факторов уже существуют реально в изучаемой совокупности, они не «ожидаемые», не «взятые с потолка». Это хорошо. Однако могут ли эти значения

факторов сочетаться в одном предприятии, системны ли эти значения? Решение данного вопроса выходит за рамки

статистики, оно требует конкретных знаний об объекте

прогнозирования.

Если, кроме количественных факторов, при многофакторном регрессионном анализе в уравнение включается и неколичественный, то применяют следующую методику: наличие неколичественного фактора у единиц совокупности обозначают единицей, его отсутствие — нулем, т.е. вводят так назы-



Число фиктивных переменных должно быть на единицу меньше числа градаций качественного (неколичественного) фактора. С помощью данного приема можно измерять влияние уровня

образования, местожительства, типа жилища и других

социальных или природных, неизмеряемых количественно факторов, изолируя их от влияния количественных факторов.