Энгельсский технологический институт (филиал) федерального
государственного бюджетного образовательного учреждения
высшего образования
«Саратовский государственный технический университет
имени Гагарина Ю.А.»

Кафедра «Оборудование и технологии обработки материалов»

РЕГРЕССИОННЫЕ МОДЕЛИ С ОДНОЙ ВХОДНОЙ ПЕРЕМЕННОЙ

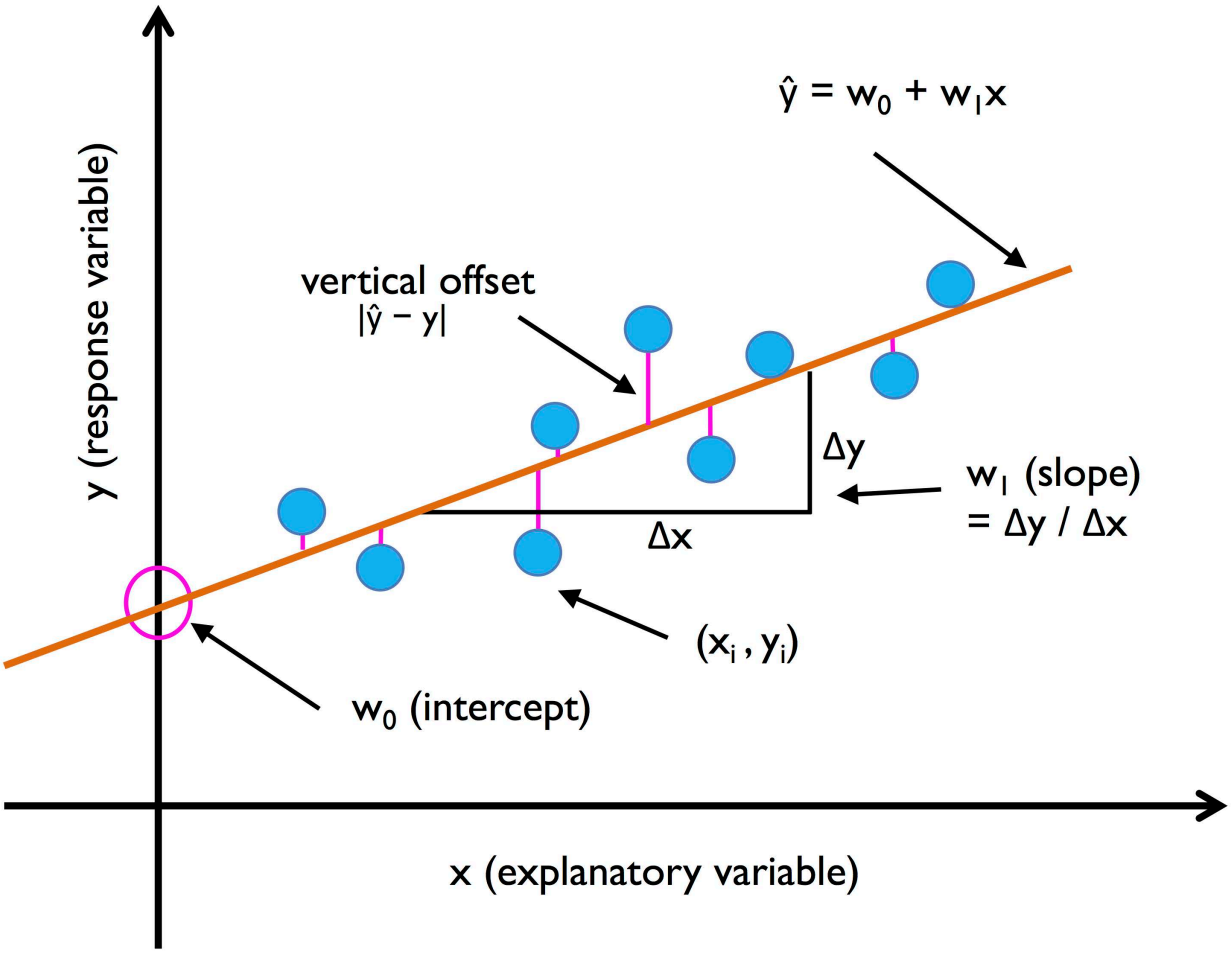
Электронное издание локального распространения

Методические указания к практической работе
по дисциплине «Математическое моделирование технологических процессов»
для студентов направлений подготовки:

15.03.05 «Конструкторско-технологическое обеспечение машиностроительных производств»

очной и заочной форм обучения

Энгельс 2024



1. Основные понятия

Технологические процессы машиностроительного производства, особенно процессы обработки резанием конструкционных материалов, очень сложны по своей физико-химической природе. До сих пор отсутствуют принятые всеми аналитические модели, точно описывающие закономерности процессов изнашивания и нагружения инструмента, тепловых процессов в зоне резания и т. д. Поэтому в технологии машиностроения очень часто используют модели, которые мы ранее обозначили как эмпи-

рические. Эмпирические модели объектов и процессов представляют собой результат обработки экспериментальных данных о поведении объекта или процесса методами математического статистического анализа. Очень часто для построения моделей объектов по результатам экспериментальных исследований используют математический аппарат регрессионного и корреляционного анализа.

Термины «*регрессия*» и «*корреляция*» были введены в широкое употребление статистиками Ф. Гальтоном и К. Пирсоном в конце XIX в. Они изучали взаимозависимости роста и массы людей разного возраста и столкнулись с необходимостью введения таких показателей указанной зависимости, которые бы отражали связь между исследуемыми характеристиками человека, но не определяли бы друг друга строго однозначно. В настоящее время «*регрессия*» и «*корреляция*» – основные понятия статистики.

Основная задача *корреляционного анализа* – выявление значимости связи между значениями различных случайных величин. Зависимость между величинами (в том числе и случайными), при которых одному значению одной величины (аргумента) отвечает одно или несколько вполне определенных значений другой величины, называется, соответственно, *однозначной* или *многозначной функциональной зависимостью* [11]. Зависимость между величинами, при которой каждому значению одной величины отвечает с соответствующей вероятностью множество возможных значений другой, называют *вероятностной* (стохастической, статистической). Примерами корреляционной связи являются зависимости между пределами прочности и текучести стали определенной марки, между погрешностями размера и погрешностью формы поверхности детали, между температурой испытания и прочностью материала и т. д.

Математический аппарат *регрессионного анализа* позволяет:

- оценить неизвестные параметры предлагаемой к исследованию регрессионной модели;
- проверить статистическую значимость параметров модели;
- проверить адекватность модели;
- оценить точность модели.

Вид регрессионной модели предлагает сам исследователь, при этом он исходит из следующего:

- физической сущности изучаемого объекта или явления;
- характера экспериментального материала;

● анализа априорной информации.

Самым простым для моделирования является объект, у которого один входной и один выходной фактор (рис. 1). Входной фактор характеризует воздействие на исследуемый объект. В технологических процессах машиностроения это могут быть температура, сила, время, геометрические параметры инструмента, характеристики обрабатываемого и инструментального материалов и т. д. Выходной фактор характеризует реакцию (отклик) объекта на воздействие входного фактора. Выходные факторы в технологических процессах машиностроения – длина пройденного инструментом пути, величина износа, напряжения, качество обработанной поверхности и т. д.

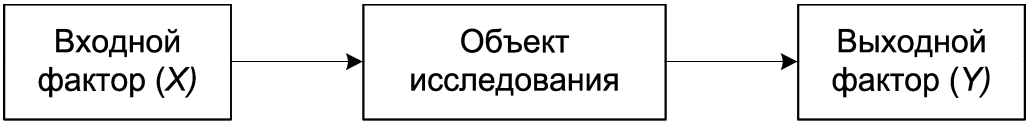


Рис. 1. Объект исследования с одним входным и одним выходным фактором

Для начала построения эмпирической модели необходимо иметь данные экспериментальных исследований объекта (в виде таблицы или графика), в которых каждому значению входного фактора (X) соответствует значение выходного фактора (Y), т. е. известна пара чисел (x_i, y_i). Пары случайных переменных (x, y) подчиняются некоторому двумерному вероятностному распределению. Общее количество пар чисел пусть равно m (рис. 2).

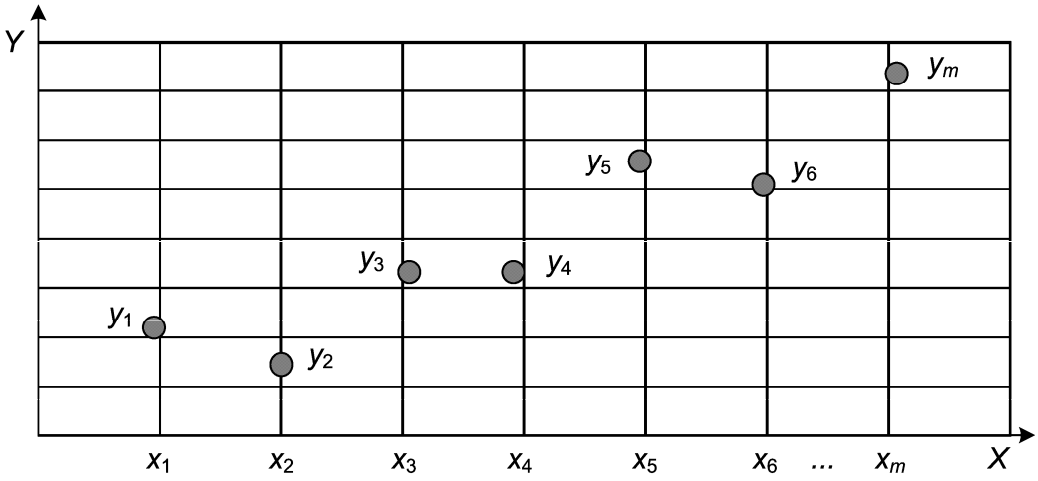


Рис. 2. Графическое отображение результатов эксперимента

Данный график называется *диаграммой рассеяния*, или *точечной диаграммой* [12]. Необходимо найти такую кривую, которая бы наилучшим образом аппроксимировала экспериментальные точки. Для удобства дальнейшего исследования объекта эта кривая должна иметь для своего описания одну единственную формулу (функцию). Если мы соединим точки на графике, то получим ломаную линию, состоящую из нескольких прямых отрезков и описываемую соответствующим количеством линейных моделей. Это крайне неудобно для исследования. Необходимо найти кривую, наилучшим образом описывающую все экспериментальные точки (рис. 3). Такую кривую называют *кривой регрессии*, или *регрессионной кривой Y* по *X*. В общем случае кривая регрессии может иметь любой вид (монотонно возрастающая, монотонно убывающая, с точками перегиба и т. д.), но она должна быть непрерывной, т. е. не должна иметь разрывов. В самом простом случае кривая регрессии имеет вид прямой линии.

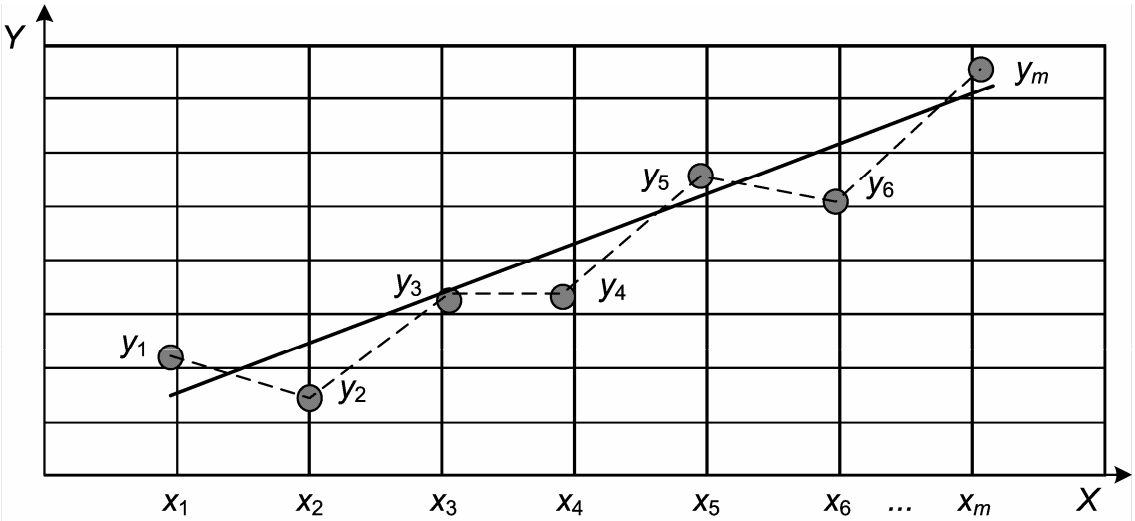


Рис. 3. Построение линии регрессии

Обычно построение моделей и исследование объекта начинают с самых простых моделей – *линейных*. Линейной модели соответствует кривая регрессии в виде *простой линии*.

Как видно из графика (см. рис. 3), всегда имеются отклонения экспериментальных точек от кривой регрессии, что вызвано влиянием других (неучтенных в модели) внешних факторов на исследуемый объект. В моделировании выходной фактор еще называют *зависимой выходной переменной*, а входной – *независимой входной переменной*. Во время исследования

объекта входной фактор всегда носит детерминированный характер, а выходной – случайный.

Выражение, которое устанавливает связь между случайной зависимой и детерминированной независимой переменными, представляет собой *уравнение регрессии*. Термин «уравнение регрессии», строго говоря, не совсем корректный [12], но общепринятый. Модель, построенная на основе уравнения регрессии, является *регрессионной моделью*. Как указывалось ранее, для получения регрессионных моделей (уравнений регрессии) используется математический аппарат регрессионного анализа.

Итак, как мы уже говорили, подбор кривой регрессии и регрессионной модели обычно начинают с простой прямой линии и, соответственно, с линейной модели. Если иметь неограниченно большое количество экспериментальных точек, то линейная регрессионная модель имеет вид [12]

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon, \tag{1}$$

$$\hat{y} = \beta_0 + \beta_1 \cdot x, \tag{2}$$

где $\hat{y}$ – значения выходной переменной, рассчитанные (предсказанные) по линейной модели;

x – значения входной переменной;

β_0 и β_1 – коэффициенты регрессии;

ε – остаток (невязка).

Определение коэффициентов регрессии осуществляется на основе *метода наименьших квадратов*. Метод наименьших квадратов применяют в тех случаях, когда случайная вариация входного фактора пренебрежительно мала по сравнению с наблюдаемым диапазоном его измерения [12], т. е. значения входной переменной считаются фиксированными. Суть метода в том, что подбираются такие β_0 и β_1, при которых сумма квадратов отклонений измеренных величин y от предсказанных $\hat{y}$ была бы минимальной.

Для пар наблюдений можно записать

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i\ . \tag{3}$$

Отклонение измеренной величины y от предсказанной $\hat{y}$

$$\varepsilon_i = y_i - \hat{y} = y_i - (\beta_0 + \beta_1)x. \tag{4}$$

Сумма квадратов отклонений выражается в виде

$$S = \sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i), \tag{5}$$

где S – функция суммы квадратов.

Подберем b_0 и b_1 так, чтобы при подстановке их вместо β_0 и β_1 значение S было минимальным из возможных. Найдем частные производные (∂):

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)^2, \tag{6}$$

$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{m} x_i (y_i - \beta_0 - \beta_1 x_i)^2. \tag{7}$$

Наименьшее значение суммы квадратов отклонений достигается в том случае, когда коэффициенты β_0 и β_1 удовлетворяют условию [11]

$$\frac{\partial S}{\partial \beta_0} = \frac{\partial S}{\partial \beta_1} = 0. \tag{8}$$

Имеющиеся экспериментальные данные в виде пар (x_i, y_i) являются лишь ограниченной выборкой из общего числа состояний исследуемого объекта. Поэтому можно определить только оценки коэффициентов β_0 и β_1, которые обозначают, соответственно, b_0 и b_1.

$$\hat{y} = b_0 + b_1 x. \tag{9}$$

Такие модели в литературе часто называют *однофакторными регрессионными моделями*. Коэффициент регрессии b_1 определяется по формуле [12]

$$b_1 = \frac{\sum_{i=1}^{m} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{m} (x_i - \overline{x})^2}, \tag{10}$$

где x_i – значение входного фактора во время эксперимента;

y_i – значение выходного фактора, соответствующее x_i;

$\overline{x}$ –среднее значение входного фактора, определяемое по формуле

$$\overline{x} = \frac{\sum_{i=1}^{m} x_i}{m}, \tag{11}$$

$\overline{y}$ –среднее значение выходного фактора, определяемое по формуле

$$\overline{y} = \frac{\sum_{i=1}^{m} y_i}{m}. \tag{12}$$

Коэффициент регрессии b_0

$$b_0 = \overline{y} - b_1\overline{x}. \tag{13}$$

Получаем

$$\hat{y} = \overline{y} + b_1(x_i - \overline{x}), \tag{14}$$

где $\overline{x}$ и $\overline{y}$ – координаты «центра тяжести» экспериментальных данных, через который обязательно проходит линия регрессии.

2. Адекватность регрессионных моделей

На определении коэффициентов регрессионной модели построение модели не заканчивается. Необходимо установить адекватность и точность предлагаемой модели. *Адекватность модели* характеризует соответствие модели экспериментальным данным и статистическую значимость уравнения регрессии. Адекватность регрессионной модели оценивается коэффициентом Фишера ($F_{\text{расч}}$)[1]

$$F_{\text{расч}} = \frac{\sum_{i=1}^{m} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}. \tag{15}$$

[1] Вывод формулы дан в кн.: *Драйпер Н.* Прикладной регрессионый анализ. М., 2007. 912 с.

Критерий Фишера представляет собой отношение суммы квадратов отклонений, обусловленных регрессией, к сумме квадратов отклонений относительно регрессии.

Формулируем нулевую гипотезу H_0: $b_1 = 0$. Расчетное значение коэффициента $F_{расч}$ необходимо сравнить с табличным значением $F_{табл}(m, \alpha)$. В этой записи m – общее количество экспериментальных наблюдений (x_i, y_i), которое влияет на количество степеней свободы при определении критерия Фишера. α – уровень значимости – вероятность, с которой мы можем отвергнуть правильную гипотезу о модели как неправильную. Обычно в моделировании используют значения $\alpha = 0{,}05$; $0{,}01$.

Если $F_{расч} > F_{табл}$, то нулевая гипотеза отвергается, модель считается адекватной, а регрессия – значимой. Если $F_{расч} < F_{табл}$, то регрессионная модель неадекватна и использовать ее для анализа и исследования объекта нельзя. В этом случае необходимо снова проанализировать априорную информацию, снова спланировать и провести эксперимент.

Значения коэффициента Фишера обычно даны в справочниках по математической статистике и теории вероятностей.

3. Точность регрессионных моделей

Для оценки *точности регрессионных моделей с одной входной* используется выборочный коэффициент корреляции Пирсона (r_{xy}), который определяется по формуле [12]

$$r_{xy} = \frac{\sum_{i=1}^{m}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{m}(y_i - \overline{y})^2}}. \qquad (16)$$

Коэффициент корреляции r_{xy} характеризует тесноту связи между выходной переменной y и входной переменной x. Область определения коэффициента корреляции r_{xy} лежит в пределах от –1 до +1 включительно. Можно выделить несколько частных случаев значения коэффициента корреляции (рис. 4)

Чем выше значение r_{xy}, тем теснее связь между выходной переменной y и входной переменной x, тем точнее, а следовательно, лучше математическая модель.

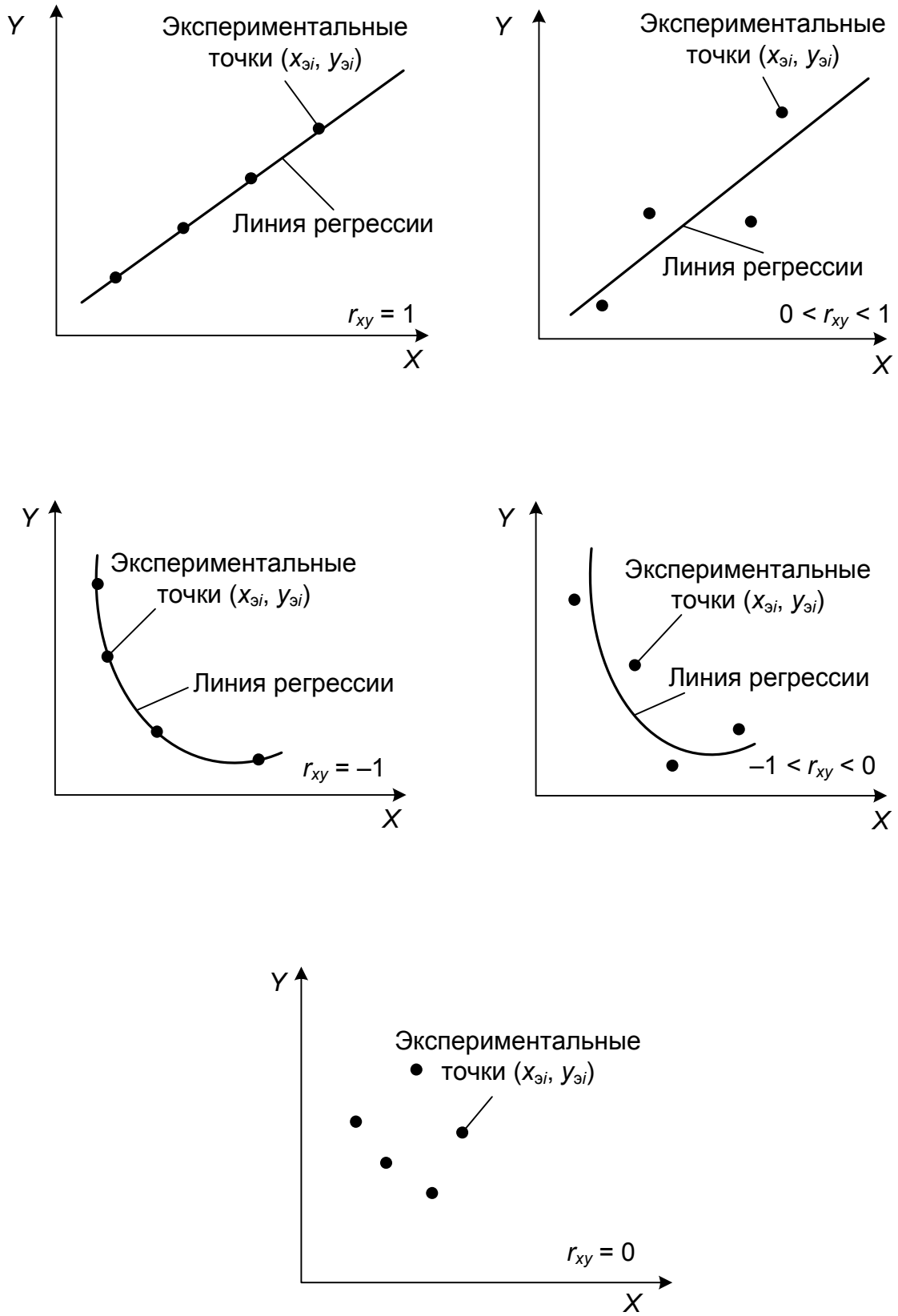


Рис. 4. Частные случаи значения коэффициента корреляции

Если модель имеет низкое значение r_{xy}, то она имеет низкую точность оценки и предсказания поведения или свойств объекта. Такую модель использовать для исследования, описания и предсказания объекта не рекомендуется. Из нескольких моделей, проанализированных во время моделирования, для исследования объекта выбирается та модель, у которой коэффициент корреляции r_{xy} имеет наибольшее значение. После расчета коэффициента корреляции производят проверку его значимости при помощи критерия Стьюдента. Коэффициент корреляции, рассчитанный для модели ($r_{расч}$), сравнивается с табличным (граничным) значением ($r_{табл}$). Если $r_{расч} > r_{табл}$, то $r_{расч}$ принимается как показатель тесноты связи и наоборот [10]. Табличные значения $r_{табл}$ можно найти в справочниках по теории вероятностей и математической статистике.

Если попытаться оценить стандартную ошибку для предсказанных значений выходного фактора $\overline{y}$, то наилучшие предсказания будут в «центре тяжести» эксперимента [12]. Чем дальше от «центра тяжести», тем менее точными будут предсказания $\overline{y}$.

4. Виды регрессионных моделей с одной входной переменной

Если в результате расчета коэффициента корреляции r_{xy} линейная модель признана недостаточно точной, переходят к исследованию более сложных моделей: *степенной* ($y = b_0x^{b_1}$), *экспоненциальной* ($y = \exp(b_0 + b_1x)$), *обратной* ($y = b_0 + \frac{b_1}{x}$ или $y = \frac{1}{b_0 + b_1x}$), *полинома* ($y = b_0 + b_1x + b_2x^2$).

Полином и обратные модели являются линейными по параметрам, поэтому для оценки их коэффициентов регрессии, корреляции и критерия адекватности можно использовать формулы (5.10), (5.13), (5.15), (5.16). Степенная и экспоненциальная модели требуют дополнительных преобразований в виде логарифмирования.

Вопросы для самоконтроля

1. Что такое линия регрессии?
2. Что такое уравнение регрессии?
3. Какие модели называются регрессионными?
4. На основе какого метода определяются коэффициенты регрессии?
5. Как определяются коэффициенты регрессии однофакторной модели?

6. Каким критерием оценивается адекватность модели с одним входным фактором?

7. Что делать, если модель оказывается неадекватной?

8. Как оценивается точность однофакторной модели?

9. Где точность предсказаний значений выходного фактора выше?

10. Какими, кроме линейной, могут быть модели с одной входной переменной?

Библиографический список

1. *Кузьмин В. В.* Математическое моделирование технологических процессов сборки и механической обработки изделий машиностроения: учебник для вузов / В. В. Кузьмин [и др.]. Москва: Высшая школа, 2008. 279 с.

2. *Ашихмин В. Н.* Введение в математическое моделирование: учебное пособие / В. Н. Ашихмин [и др.]; под ред. П. В. Трусова. Москва: ЛОГОС, 2005. 440 с.

3. *Советов Б. Я.* Моделирование систем: учебник для вузов / Б. Я. Советов, С. А. Яковлев. 3-е изд., перераб и доп. Москва: Высшая школа, 2001. 343 с.

4. *Зобнин Б. Б.* Моделирование систем: конспект лекций / Б. Б. Зобнин. Екатеринбург: Изд-во УГГГА, 2001. 129 с.

5. *Дьяконов В. П.* Новые информационные технологии: учебное пособие / В. П. Дьяконов [и др.]; под ред. В. П. Дьяконова. Москва: СОЛОН-Пресс, 2005. 640 с.

6. *Саблина Н. Г.* Информационные технологии: конспект лекций: в 2 частях / Н. Г. Саблина, Г. М. Черногородова. Екатеринбург: Изд-во УГТУ – УПИ, 2001. Ч. 2. 119 с.

7. *Дулов В. Г.* Математическое моделирование в современном естествознании: учебное пособие / В. Г. Дулов, В. А. Цибаров; под ред. В. Г. Дулова. Санкт-Петербург: Изд-во С.-Петерб. ун-та, 2001. 244 с.

8. *Зарубин В. С.* Математическое моделирование в технике: учебник для вузов / В. С. Зарубин [и др.]; под ред. В. С. Зарубина. Москва: Изд-во МГТУ им. Н. Э. Баумана, 2001. 496 с.

9. *Корн Г.* Справочник по математике для научных работников и инженеров / Г. Корн, Т. Корн. Москва: Наука, 1972. 830 с.

10. *Спирин Н. А.* Методы планирования и обработки результатов инженерного эксперимента: учебное пособие / Н. А. Спирин [и др.]; под ред. Н. А. Спирина; ГОУ ВПО УГТУ – УПИ. Екатеринбург, 2003. 260 с.

11. *Рогов В. А.* Методика и практика технических экспериментов: учебное пособие / В. А. Рогов. Москва: Академия, 2005. 288 с.

12. *Дрейпер Н.* Прикладной регрессионный анализ: перевод с английского / Н. Дрейпер, Г. Смит. 3-е изд. Москва: Вильямс, 2007. 912 с.

13. *Адлер Ю. П.* Теория эксперимента: прошле, настоящее, будущее / Ю. П. Адлер, Ю. В. Грановский, Е. В. Макарова. Москва: Знание, 1982. 64 с.

14. *Цирлин А. М.* Оптимальное управление технологическими процессами / А. М. Цирлин. Москва: Энергопромиздат, 1986. 400 с.

15. *Ногин В. Ю.* Основы теории оптимизации / В. Ю. Ногин, И. О. Протодьяконов, И. И. Евлампиев. Москва: Высшая школа, 1986. 384 с.