# Математическое моделирование

Статистическая обработка экспериментальных данных

# РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессионный анализ - это метод установления аналитического выражения стохастической зависимости между исследуемыми признаками. Уравнение регрессии показывает, как в среднем изменяется *y* при изменении любого из $x_i$, и имеет вид:

$$y = f(x_1, x_2 \cdots x_n)$$

где *y* - зависимая переменная (она всегда одна);

$x_i$ - независимые переменные (факторы) (их может быть несколько).

Наиболее распространенным методом поиска коэффициентов уравнений регрессии является метод наименьших квадратов, разработанный Лежандром и Гауссом.

Метод заключается в поиске минимума функции

$$F(b_1, b_2 \cdots b_n) = \sum_i (\Delta y_i)^2 \xrightarrow[b_1, b_2 \cdots b_n]{} min$$

# РЕГРЕССИОННЫЙ АНАЛИЗ

где

$$\Delta y_i = y_i - \mathrm{f}(x_i, b_1, b_2 \cdots b_n)$$

называются *остаточными погрешностями*.

В точке минимума частные производные функции остаточных погрешностей по каждому параметру должны обращаться в нуль, что приводит к системе уравнений

$$\frac{\partial \Phi(b_1, b_2 \cdots b_n)}{\partial b_j} = \sum_i (y_i - \mathrm{f}(x_i, b_1,\ b_2 \cdots b_n))^2 \frac{\partial f(x_i, b_1,\ b_2 \cdots b_n)}{\partial b_j} = 0$$

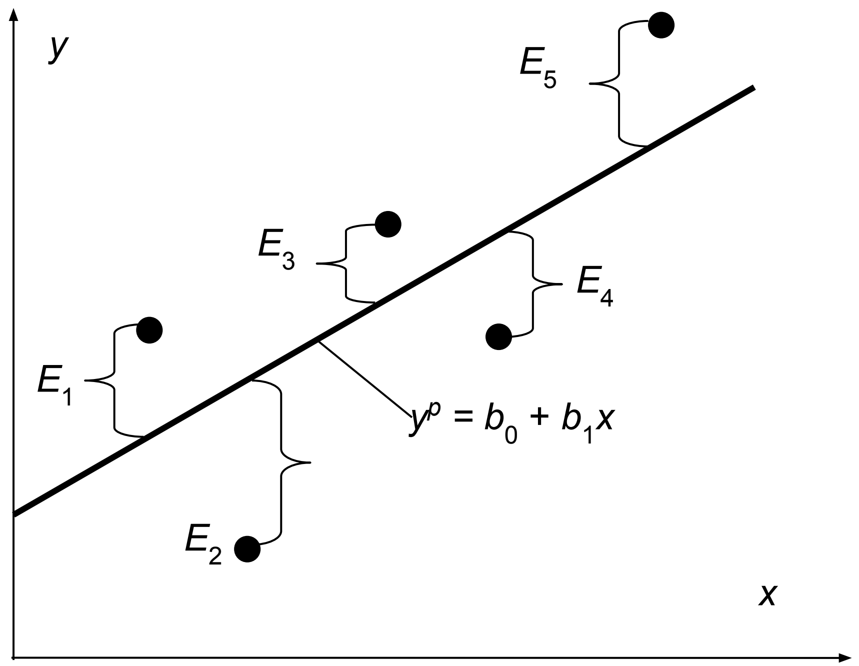
# РЕГРЕССИОННЫЙ АНАЛИЗ

Допустим, что модель представляет собой прямую линию:

$$y^{\mathrm{p}} = b_0 + b_1 x$$

где $y^{\mathrm{p}}$ – величина, предсказываемая регрессионной моделью.

Требуется получить такие значения коэффициентов $b_0$, $b_1$, при которых сумма квадратов ошибок является минимальной. Ошибка $E$ для каждой точки определяется как расстояние по вертикали от этой точки до прямой линии (рис).



# РЕГРЕССИОННЫЙ АНАЛИЗ

Обозначим

$$\begin{cases} y^{\text{p}}{}_1 = b_0 + b_1 x_1 \\ y^{\text{p}}{}_2 = b_0 + b_1 x_2 \\ \qquad \cdots \\ y^{\text{p}}{}_n = b_0 + b_1 x_n \end{cases}$$

тогда ошибки будут выражаться в виде

$$\begin{cases} \mathrm{E}_1 = y^{\text{p}}{}_1 - y_1 = b_0 + b_1 x_1 - y_1 \\ \mathrm{E}_2 = y^{\text{p}}{}_2 - y_2 = b_0 + b_1 x_2 - y_2 \\ \qquad \cdots \\ \mathrm{E}_{\mathrm{n}} = y^{\text{p}}{}_n - y_n = b_0 + b_1 x_n - y_n \end{cases}$$

# РЕГРЕССИОННЫЙ АНАЛИЗ

Функция ошибки *F* определяется выражением $F = \mathrm{E}_1{}^2 + \mathrm{E}_2{}^2 \cdots \mathrm{E}_n{}^2$ или

$$F = \sum_{i=1}^{n} (b_0 + b_1 - y_i)^2$$

Для получения таких значений $b_0$ и $b_1$, при которых функция *F* является минимальной, применяются обычные методы математического анализа. Условиями минимума являются $\frac{\partial F}{\partial b_0} = 0, \frac{\partial F}{\partial b_1} = 0$ .

Дифференцируя *F*, получаем

$$\frac{\partial F}{\partial b_0} = \frac{\partial}{\partial b_0} \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 = \sum_{i=1}^{n} 2(b_0 + b_1 x_i - y_i) = 2(nb_0 + b_1 \sum x_i - \sum y_i) = 0,$$

$$\frac{\partial F}{\partial b_1} = \frac{\partial}{\partial b_1} \sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2 = \sum_{i=1}^{n} 2x_i (b_0 + b_1 x_i - y_i) = 2b_0 \sum_{i=1}^{n} x_i + 2b_1 \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} y_i x_i = 0,$$

# РЕГРЕССИОННЫЙ АНАЛИЗ

$$\begin{cases} nb_0 + b_1 \sum_i x_i = \sum_i y_i \\ b_0 \sum_i x_i + b_1 \sum_i {x_i}^2 = \sum_i x_i\, y_i \end{cases}$$

Решая систему двух линейных алгебраических уравнений, можно получить значения $b_0$ и $b_1$.

# РЕГРЕССИОННЫЙ АНАЛИЗ

$$b_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad b_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

**Обычно мерой ошибки регрессионной модели служит стандартное (среднеквадратичное) отклонение *S (остаточная дисперсиия)*, определяемое по формуле**

$$S = \sqrt{\frac{\sum_{i=1}^{n} E_i^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} (b_0 + b_1 x_i - y_i)^2}{n-2}}.$$

# РЕГРЕССИОННЫЙ АНАЛИЗ

Линеаризация нелинейных зависимостей производится путем соответствующей замены переменных. Примеры такой замены приведены в таблице.

| Исходная функция | Замена переменных | Новая функция |
|---|---|---|
| $y = Ax^n$ | $X = x^n$ , $a = A$ | $y = aX$ |
| $y = Ax^n$ | $Y = \ln y$, $X = \ln x$,<br>$a = n$, $b = \ln A$ | $Y = aX + b$ |
| $y = Ae^{ax}$ | $Y = \ln y$, $b = \ln A$ | $Y = ax + b$ |
| $y = ax^n + b$ | $X = xn$ | $y = aX + b$ |
| $y = ax^n + bx^m$ | $Y = yx^{-m}$, $X = xn^{-m}$ | $Y = aX + b$ |
| $y = a\sin x + b\cos x$ | $Y = y/\cos x$ , $X = \mathrm{tg}\, x$ | $Y = aX + b$ |

# РЕГРЕССИОННЫЙ АНАЛИЗ

Проверка соответствия установленной зависимости экспериментальному материалу (*проверка адекватности*) включает в себя три этапа.

1. Ищется остаточная дисперсия, или *дисперсия адекватности*

$$S_{\text{ад}}^2 = \frac{1}{f_{\text{ад}}} \sum_{i=1}^{n} \left( y_i^{\text{р}} - y_i \right)^2$$

$$f_{\text{ад}} = (n - m)$$

где

– количество степеней свободы, равное разности количества опытов *n* и количества коэффициентов в уравнении регрессии *m*. Дисперсия адекватности будет тем меньше, чем лучше совпадают расчетные значения параметра  с экспериментальными данными .

# РЕГРЕССИОННЫЙ АНАЛИЗ

2. Определяется *дисперсия воспроизводимости*, показывающая точность определения параметра в опыте. В случае, если для каждого сочетания уровней факторов проводилось несколько параллельных опытов, ищутся дисперсии $S_j^2$ для каждой группы опытов, проверяется их однородность и затем определяется средневзвешенная дисперсия $S_{\bar{y}}^2$, которая и принимается в качестве дисперсии воспроизводимости $S_{\text{в}}^2$.

$$S_{\bar{y}}^2 = \frac{1}{m}\sum_{j=1}^{m} S_j^2 = S_{\text{в}}^2$$

Если параллельные опыты не проводятся, то в качестве средневзвешенной дисперсии принимается

$$S_{\text{в}}^2 = \left(\frac{\Delta Y_{\text{пред}}}{2}\right)^2$$

где – $\Delta Y_{\text{пред}}$ предельная абсолютная погрешность измерения прибора

# РЕГРЕССИОННЫЙ АНАЛИЗ

**3. Проверяется однородность дисперсий адекватности и воспроизводимости по критерию Фишера**

$$F = \frac{S_{\text{ад}}^2}{S_{\text{в}}^2} < F_\alpha\left(f_{\text{ад}}, f_{\text{в}}\right)$$

**где** $f_{\text{в}} = \sum_{u=1}^{n}\left(n_u^{\text{пар}} - 1\right)$

**количество степеней свободы дисперсии воспроизводимости;**

$n_u^{\text{пар}}$ **– количество параллельных опытов для *u*-го сочетания уровней факторов.**

**Если расчетное значение критерия Фишера окажется меньше табличного, то полученное уравнение регрессии адекватно эксперименту с уровнем значимости *α*.**

# ИСПОЛЬЗУЕМАЯ ЛИТЕРАТУРА

1. Горев М.В. Экспериментальные методы исследований. Как писать статьи и отчеты по лабораторным, практическим и курсовым работам – Красноярск: Сибирский федеральный универститет, 2007. – 128 с. Код. Доступа https://www.twirpx.com/file/816456/