СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

(активный эксперимент)

Регрессионный анализ состоит из трех основных этапов:

1.Оценка дисперсии воспроизводимости (оценка ошибки опыта)

A) Определяется среднее по результатам опыта

$$\boxed{\bar{y}_i = \frac{\sum_{u=1}^{m} y_{iu}}{m}; i = 1,\ldots N}$$

m – число параллельных опытов.

B) Выборочные (построчные) дисперсии.

$$\boxed{S_i^{\ 2} = \frac{\sum_{u=1}^{m} \left(y_{iu} - \bar{y}_i\right)^2}{m-1};}$$

C) $\sum_{i=1}^{N} S_i^{\ 2};$

D) $$\boxed{G = \frac{S\max^2}{\sum S^2_{\ i}};}$$

$S^2 \max$ - максимальное значение выборочной дисперсии.

Проверяем дисперсии на однородность, если $G<G_{табл.}(q,f_1,f_2)$, то дисперсия однородна. Если дисперсия не однородна, увеличивают число параллельных опытов и вновь проверяют дисперсию на однородность.

E) Рассчитывается дисперсия воспроизводимости.

$$\boxed{S_{воспр.}^{\ \ 2} = \frac{\sum S^2_{\ i}}{N};}$$

2. *Оценка значимости коэффициентов проводится по критерию Стьюдента*

$$t_{b_i} = \frac{|b_i|}{S_{b_i}}$$

$|b_i|$ - абсолютное значение коэффициента регрессии.

Sb_i - средне квадратичное отклонение i – того коэффициента.

$$\boxed{S_{b_i}^2 = \frac{S_{воспр.}^2}{N};}$$

Если $t_{b_i} > t_{табл.}(q, f), f = N(m-1)$, то коэффициент значим. Если нет, то коэффициент приравнивается к 0 и из уравнения исключается.

Следовательно, фактор, при котором стоит этот коэффициент на данный процесс влияет незначительно.

$$F = \frac{S_{ост.}^2}{S_{воспр.}^2};$$

$$S_{ост}^2 = \frac{\sum\left(\bar{y}_i^{э} - \hat{y}_i\right)^2}{N-l}$$

$$l = n+1$$

3. *Проверка модели на адекватность.*

Если $F<F_{табл.}(q,f_1,f_2)$, то линейное уравнение регрессии адекватно описывает процесс.

$f_1=N-n-1$

$f_2=N(m-1)$

<u>ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ</u>

ПФЭ является эффективным средством построения математической модели исследуемого объекта. Однако одним из недостатков ПФЭ является то, что с увеличением числа факторов резко возрастает количество опытов. Например:

2^7=128 опытов

2^{15}=32768 опытов

Для сокращения количества опытов (с целью экономии времени) пользуются <u>*дробными репликами*</u> от ПФЭ или <u>*дробным факторным экспериментом.*</u>

Идея ДФЭ заключается в том, чтобы сократить число опытов ПФЭ, но при этом матрица планирования должна сохранить свойства ортогональности.

Дробным факторным экспериментом называется система опытов, представляющая собой часть ПФЭ, позволяющая рассчитать коэффициенты уравнения регрессии и сократить количество опытов.

Для построения статистической модели процесса используется определенная часть ПФЭ: $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ и т.д. Эти системы (планы) опытов называются ***дробными репликами.***

Для того, чтобы дробная реплика представляла собой ортогональный план, в качестве реплики берут ближайший ПФЭ. При этом число опытов должно быть больше или равно числу неизвестных коэффициентов в уравнении регрессии.

Необходимо исследовать влияние на результат ХТП трех факторов и получить его математическое описание в виде линейного уравнения .

Матрица планирования для трех факторов: в ПФЭ $N=2^3=8$

N	X_0	X_1	X_2	x1*x2 (X_3)
1	+	-	-	- +
2	+	+	-	- -
3	+	-	+	- -
4	+	+	+	- +
5	+	-	-	+ +
6	+	+	-	+ -
7	+	-	+	+ -
8	+	+	+	+ +

Допустим, по каким-то причинам нам необходимо сократить число опытов. При этом свойства матрицы планирования должны быть сохранены, а число опытов N не должно быть меньше 4 (число коэффициентов линейной модели для трех факторов).

Для решения этой задачи возьмем ближайший ПФЭ 2^2 и предположим, что взаимодействие между факторами x_1 и x_2 в ПФЭ равно нулю. Поэтому в качестве плана для x3 новой матрицы используем взаимодействие $x_1 \cdot x_2$

Получим дробную реплику (в частности, полуреплику 1/2) 2^{3-1} от ПФЭ 2^3, которая сохраняет все свойства матрицы планирования

N	x_0	x_1	x_2	x_3 (x_1x_2)
1	+	+	+	+
2	+	–	+	–
3	+	+	–	–
4	+	–	–	+

Число опытов для ДФЭ определяется по формуле:

$$N=2^{n-p}$$

n – общее число факторов;

p – число факторов, приравненных к произведению.

Если n=3, $x_3=x_1x_2$, то $N = 2^{n-1} = 2^{3-1} = 2^2 = 4$.

Применение ДФЭ всегда связано со смешиванием, т.е. совместным оцениванием нескольких теоретических коэффициентов модели. Если коэффициенты регрессии при парных произведениях не равны 0, то найденные нами коэффициенты будут <u>смешанными оценками</u> для генеральных коэффициентов.

$$b_0 = \beta_0 + \beta_{1,2,3};$$
$$b_1 = \beta_1 + \beta_{23};$$
$$b_2 = \beta_2 + \beta_{13};$$
$$b_3 = \beta_3 + \beta_{12},$$

β_i – истинные коэффициенты;

b_1 – оценки коэффициентов, вычисленных по данным выборки.

По данному плану (ДФЭ) можно вычислить коэффициенты b_0, b_1, b_2, b_3, однако, они смешаны с парными и тройными взаимодействиями.

Таким образом, сокращение числа опытов приводит к корреляции между столбцами матрицы ДФЭ. В результате мы не можем раздельно оценить эффекты факторов и эффективности взаимодействий. Получаем совместные (смешанные) оценки.

<u>РАЗРЕШАЮЩАЯ СПОСОБНОСТЬ ДРОБНЫХ РЕПЛИК.</u>

При использовании ДФЭ возникает смешение эффектов, т.е. совместная оценка коэффициентов регрессии.

Введем понятие <u>генерирующего соотношения</u>, при помощи которого задается дробная реплика. <u>Генерирующим</u> называется соотношение, показывающее, какие взаимодействия заменены новыми факторами.

Например, ДФЭ типа 2^{3-1} может быть представлен двумя полурепликами, каждая из которых задается при помощи следующих генерирующих соотношений:

$x_3=x_1x_2$ или $x_3=-x_1x_2$

Верхняя часть ПФЭ 2^2?

Нижняя часть ПФЭ 2^3?

Умножим обе части генерирующего соотношения на x_3 $x_3^2 = x_1 \cdot x_2 \cdot x_3; x_3^2 = -x_1x_2x_3$

Известно, что $x_i^2 = 1$, следовательно, получим: $1=x_1x_2x_3$; $1=-x_1x_2x_3$;

Эти соотношения называются <u>определяющим контрастом</u>

Зная определяющий контраст можно найти смешенные эффекты, т.е. определить разрешающую способность дробной реплики (найти какие из коэффициентов являются несмешанными оценками). Для этого определяющий контраст последовательно умножим на каждый фактор, получим:

$$x_1^2 = 1;$$

$x_1=x_2x_3$; $x_2=x_1x_3$; $x_3=x_1x_2$;

$x_1=-x_2x_3$; $x_2=-x_1x_3$; $x_3=-x_1x_2$;

Это означает, что например, коэффициент b_1 – есть смешанный эффект для β_1 и β_{23}.

$b_1 \to \beta_1+\beta_{23}$; $b_1=\beta_1-\beta_{23}$;

$b_2 \to \beta_2+\beta_{13}$; $b_2=\beta_2-\beta_{13}$;

$b_3 \to \beta_3+\beta_{12}$ $b_3=\beta_3-\beta_{12}$;

Следовательно, резкое сокращение числа опытов приводит к смешанным оценкам коэффициентов, а, следовательно, к снижению точности вычислений.

Эффективность системы смешивания факторов и взаимодействий факторов определяется разрешающей способностью матрицы. Она будет максимальной, если линейные эффекты смешаны с произведениями наибольшего количества факторов.

Например, для планирования 2^{4-1} возможны 8 вариантов решений:

$x_4=x_1x_2$	$x_4=-x_1x_2$	$x_4=x_2x_3$	$x_4=-x_2x_3$
$x_4=x_1x_3$	$x_4=-x_1x_3$	$x_4=x_1x_2x_3$	$x_4=-x_1x_2x_3$

Наибольшая разрешающая способность у 7- ой и 8- ой реплик, т.к. тройные взаимодействия менее важны, чем двойные (наличие взаимодействия между тремя факторами менее вероятно, поэтому мы допускаем меньшую ошибку). В реальных задачах тройные взаимодействия бывают = 0 значительно чаще, чем двойные.

Поэтому, для данного случая **генерирующими** соотношениями могут быть: $\mathbf{x_4=x_1x_2x_3}$ $\mathbf{x_4=-x_1x_2x_3}$ (тройное взаимодействие приравнивается x_4).

Определяющие контрасты: (x_4) $1=x_1x_2x_3x_4$; $1=-x_1x_2x_3x_4$

$x_1=x_2x_3x_4$	$b_1=\beta_1+\beta_{234}$
$x_2=x_1x_3x_4$	$b_2= \beta_2+\beta_{134}$
$x_3=x_2x_4x_1$	$b_3=\beta_3+\beta_{124}$
$x_4=x_1x_2x_3$	$b_4=\beta_4+\beta_{123}$
$x_1x_2=x_3x_4$	$b_{12}=\beta_{12}+\beta_{34}$
$x_1x_3=x_2x_4$	$b_{13}=\beta_{13}+\beta_{24}$
$x_1x_4=x_2x_3$	$b_{14}=\beta_{14}+\beta_{23}$
$x_2x_3=x_1x_4$	$b_{23}=\beta_{23}+\beta_{14}$

Если известно, что все тройные взаимодействия незначимы (=0), то можно найти раздельные оценки для линейных эффектов и совместные для парных произведений.

Чем меньшую реплику мы берем от ПФЭ, тем сильнее смешанные эффекты.

ПФЭ (4 опыта)

от ПФЭ (2 опыта).

Расчет коэффициентов регрессии и исследование полученного уравнения регрессии при использовании ДФЭ аналогичны методике ПФЭ.