Обработка результатов совместных измерений.

Вопросы:

1. Методика регрессионного анализа.

2. Проверка статистической гипотезы об адекватности модели.

Совместные измерения представляют собой производимые одновременно измерения двух или нескольких, как правило, неодноименных величин для нахождения зависимости между ними.

Этот вид измерений находит широкое применение в научных, технических и метрологических измерениях.

Совместные измерения применяются в метрологической практике при экспериментальном определении градуировочных характеристик средств измерений, в том числе различных преобразователей.

Определение градуировочной характеристики средства измерения называется градуировкой средства измерения.

Градуировочная характеристика средства измерения представляет собой зависимость между значениями величин на входе и выходе средств измерений. Она может быть представлена в виде таблицы, графика или формулы (т. е. в аналитическом виде).

Наиболее универсальной формой градуировочной характеристики является ее представление в виде формулы, которую удобно использовать при автоматизированных испытаниях с применением ЭВМ.

Каждому измерительному прибору или преобразователю соответствуют собственная индивидуальная зависимость между входной величиной X и выходной Y, которая в общем случае зависит также и от времени t. Функциональная зависимость $y = f\ (x,t)$ представляет собой функцию преобразования измерительного прибора (или преобразователя) и является градуировочной характеристикой.

При градуировке выполняют совместные измерения входных и выходных величин. Если число точек измерения n, то получают набор результатов измерений $(x_i;y_i)$, $i = 1 \ldots n$, по которым определяют градуировочную характеристику.

В каждой исследуемой точке измерения проводятся многократно (при прямом и обратном направлении изменения входной величины).

Наиболее предпочтительной градуировочной характеристикой является линейного вида:

$$Y = \alpha + \beta \cdot X ,$$

где α – константа (свободный член);

β – коэффициенты, которые определяют по экспериментальным данным при градуировке средства измерения методом регрессионного анализа.

- В регрессионном анализе для определения коэффициентов применяют метод наименьших квадратов (МНК), который предполагает, что выполнены два основных требования:

1) значения входных величин x_i известны точно;

2) результаты измеренных выходных величин y_i содержат независимые случайные погрешности, которые распределены по нормальному закону.

Необходимо специально проверить справедливость выполнения условия. Резко выделяющиеся значения (промахи) должны быть исключены. Для этого применяют рассмотренные выше критерии проверки статических гипотез.

Для определения коэффициентов a и b в уравнении регрессии используют регрессионный анализ:

$$\bar{y} = a + b \cdot x ,$$

где $\bar{y}$ – линия регрессии (функция отклика).

1. Методика регрессионного анализа.

- Рассмотрим методику регрессионного анализа для пассивного эксперимента, когда эксперимент заранее не планируется.

Уравнению $Y = \alpha + в \cdot X$ соответствует парная регрессия, коэффициенты которой определяют по формулам:

$$b = \frac{\left(\sum_{i=1}^{n} x_i \cdot y_i - \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i\right)^2\right)}$$

$$a = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} y_i - b \cdot \sum_{i=1}^{n} x_i\right)$$

Дисперсия $\overline{Y}$ будет складываться из двух компонентов – дисперсии параметра a и дисперсии параметра b .

Верхняя y_h и нижняя y_l границы для $\overline{Y}$ имеют вид:

$$y_{ih} = \overline{Y}_i + t_q \cdot S_{\overline{y}_i}, \qquad y_{il} = \overline{Y}_i - t_q \cdot S_{\overline{y}_i},$$

- где t_q – коэффициент Стьюдента.

Среднее квадратическое отклонение отклика y вычисляют по формуле:

$$S_{\bar{y}_i} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}_i\right)^2}{(n-2)} \cdot \left[\frac{1}{n} + \frac{(x_i - \overline{X})^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}\right]},$$

где $x = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

Для определения доверительной области с учетом отклонений отдельных измерений необходимо вычислить среднее квадратическое отклонение по формуле:

$$S_{\bar{y}_i} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \overline{Y}_i\right)^2}{(n-2)} \cdot \left[1 + \frac{1}{n} + \frac{(x_i - \overline{X})^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}\right]}$$

- Тогда доверительные границы будут равны:

$$y_{ih} = \overline{Y}_i + t_q \cdot S_{y_i}, \qquad y_{il} = \overline{Y}_i - t_q \cdot S_{y_i}$$

Если возникает необходимость проверки статической гипотезы о равенстве двух уравнений регрессии, то такая проверка включает последовательную проверку справедливости трех статических гипотез:

1) об остаточных дисперсиях;

2) о значениях коэффициентов регрессии *b*;

3) о значениях коэффициентов (константа) *a*.

В рассмотренной парной регрессии значения y зависят от значений только одной переменной *x* . Однако в общем случае y может зависеть от нескольких переменных x_1, x_2, …, x_n. Это так называемый случай множественной регрессии.

Оценка параметров регрессии обычно сопровождается расчетом дополнительной характеристики, называемой коэффициентом корреляции.

- Выборочный коэффициент корреляции представляет собой эмпирическую (т. е. определенную по экспериментальным результатам) меру линейной зависимости между *x* и *y*.

В математической статистике степень коррелируемости переменных (*n* пар случайных величин x_i, y_i, i=1, 2, 3, …, *n*), которую оценивают выборочным коэффициентом корреляции Пирсона:

$$r_b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]}}$$

Если $r_b > 0$, то при увеличении *x* возрастает *y* , при $r_b < 0$, с ростом *x* , *y* – убывает. Принято считать, что при выполнении условия $0{,}75 < |r_b| < 0{,}95$, существует сильная связь, а при $0{,}95 < |r_b| \leq 1$ - функциональная зависимость.

Для небольших значений n (так называемая малая выборка) коэффициент корреляции r должен быть скорректирован:

$$r_b' = r_b \cdot \left[1 + \frac{1 - r_b^2}{2 \cdot (n-3)}\right]$$

Выборочный коэффициент r_b (или r_b') должен быть проверен на существенность, т. е. значимо ли он отличается от нуля.

Для проверки статической гипотезы о существенности (значимости) корреляции между исследуемыми величинами X и Y и построения доверительных интервалов для коэффициента корреляции используют преобразование Фишера:

$$U = \frac{1}{2} \cdot \ln \frac{1+r}{1-r},$$

которое аппроксимируется нормальным законом с дисперсией:

$$\sigma_U^2 = (n-3)^{-1}.$$

- Если $|U| > Z_{1-q/2} \cdot \sigma_U$, то коэффициент корреляции существенно отличается от нуля.

Приняв, что справедлив нормальный закон (для доверительной вероятности *P* = 0,95), находят верхнюю U_h и нижнюю U_l границы доверительного интервала:

$$U_h = U + 1{,}96 \cdot \sigma_U; \qquad U_l = U - 1{,}96 \cdot \sigma_U .$$

Если число измерений $n > 50$, то при доверительной вероятности 0,95 доверительные границы выборочного значения коэффициента корреляции имеют вид:

$$r_h = r_b + 1{,}96 \cdot \frac{1-r_b^2}{\sqrt{n}}; \qquad r_h = r_b - 1{,}96 \cdot \frac{1-r_b^2}{\sqrt{n}}$$

С помощью преобразования $U = \frac{1}{2} \cdot \ln \frac{1+r}{1-r}$ можно установить равенство между собой двух выборочных коэффициентов корреляции, используя, например, статический критерий *T*.

- При выбранном уровне значимости q требуется проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции (при конкурирующей гипотезе: выборочный коэффициент корреляции r_b отличен от нуля).

Если нулевая гипотеза отвергается, то это означает, что выборочный коэффициент корреляции значимо отличается от нуля, а X и Y коррелированы, т. е. связаны линейной зависимостью.

В качестве критерия проверки нулевой гипотезы принимают случайную величину:

$$T = \frac{r_b \cdot \sqrt{n-2}}{\sqrt{1 - r_b}}$$

Величина T при справедливости нулевой гипотезы имеет распределение Стьюдента с $k = n - 2$ степенями свободы.

2. Проверка статистической гипотезы об адекватности модели.

- Целью проверки адекватности математической (регрессионной) модели является подтверждение того, что данная модель правильно описывает исследуемый процесс. Для этого определяются погрешности математической модели и экспериментальных данных. Если погрешности модели превышают погрешности экспериментальных данных, то гипотеза об адекватности математической модели отклоняется.

$$F = \frac{S_1^2}{S_2^2}$$

$$S_{\text{ср}}^2 = \sum_{i=1}^{m} \frac{S_i^2}{m}$$

$$G < G_{1-q(f;m)}$$

- Для проверки используют критерий Фишера $F = \frac{S_1^2}{S_2^2}$, в числитель которого записывают дисперсию адекватности $S_{\text{ад}}^2$, а в знаменатель – усредненную дисперсию $S_{\text{ср}}^2$, которая вычисляется по формуле $S_{\text{ср}}^2 = \sum_{i=1}^{m} \frac{S_i^2}{m}$ при условии подтверждения однородности дисперсий по $G < G_{1-q(f;m)}$.

Дисперсию адекватности определяют по формуле:

$$S_{\text{ад}}^2 = \sum_{i=1}^{n} \frac{(\overline{Y}_i - \overline{Y})^2}{(n-c)}$$

где c – число коэффициентов уравнения регрессии;

$\overline{Y}_i$ – среднее арифметическое значение экспериментальных данных;

$\overline{Y}$ – значение функции отклика, вычисленное по уравнению регрессии.