Обработка многократно измеренных величин

Анализируемый случай:

1. Закон распределения погрешностей известен, или неизвестен. Параметрический – непараметрический подход
2. Случайность – отсутствие ярко выраженной закономерности (для геодезии значимые систематич. влияния)
3. Однородность – в выборке нет грубых измерений и все части выборки имеют примерно одинаковую оценку сдвига и масштаба
4. Независимость – элементы в ряде достаточно независимы (не коррелированы) между собой

Обработка многократно измеренных величин

Исследование на нормальность:

1. Предварительные исследования (грубость, систематика – условия Ляпунова – основн. мат. условия)
2. Графические исследования – гистограмма, вероятностная бумага
3. Приближенные исследования (форма –ассиметрия, эксцесс – важно для тестирования)
4. Основные критерии на основе проверки гипотез
   - критерий χ_2-***Пирсона***
   - критерий ***Мизеса–Крамера– Смирнова***

Обработка многократно измеренных величин

Суть критерия Пирсона:

Подсчет разности практических и теоретических относительный частот, попавших в соответствующий интервал. Вывод по степени различия

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(n_j - n \cdot p_j\right)^2}{n \cdot p_j} \leq \chi_t^2(p,\ k)$$

Суть критерия Мизеса-Крамера-Смирнова:

Подсчет взвешенной разности практических и теоретических накопленных частот. Вывод по степени различия

$$n \cdot \omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(P(x) - W(x)\right)^2 \leq \omega_t^2$$

Обработка многократно измеренных величин

Случайность – параметрический подход

Неслучайность -линейный тренд - систематика.

Критерий коэффициентов регрессии. Модель

$$(y_m)_i = y_i + v_i = a \cdot i + b$$

Суть – значимость a. Выявление по МНК.

Функция качества (целевая функция) Φ

$$\Phi = \left[v^2 \right] = \left[\left(a \cdot i + b - y_i \right)^2 \right]$$

Для нахождения a и b – производные от Φ и к нулю

$$\begin{cases} \dfrac{\partial \Phi}{\partial a} = 2 \cdot \left[\left(a \cdot i + b - y \right) \cdot i \right] \Rightarrow \left[i^2 \right] a + [i] b - \left[y \cdot i \right] = 0 \\ \dfrac{\partial \Phi}{\partial b} = 2 \cdot \left[\left(a \cdot i + b - y \right) \cdot 1 \right] \Rightarrow [i] a + n \cdot b - \left[y \right] = 0 \end{cases}$$

Обработка многократно измеренных величин

Решить нормальные уравнения относительно *a* и *b*

$$\begin{cases} [i^2]a + [i]b - [x \cdot i] = 0 \\ [i]a + n \cdot b - [x] = 0 \end{cases} \left| \begin{array}{l} n_{11}a + n_{12}b = c_1 \\ n_{21}a + n_{22}b = c_2 \end{array} \right.$$

В матричном виде

$$N \cdot k = c \Rightarrow \begin{bmatrix} [i^2] & [i] \\ [i] & n \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} [x \cdot i] \\ [x] \end{bmatrix}$$

Лучшая вычислительная схема:

- Составляем матрицу планы *A* из коэффициентов при неизвестных *a* и *b*

$$A = \begin{bmatrix} 1\,1 \\ 2\,1 \\ \ldots 1 \\ n\,1 \end{bmatrix}$$

$$(y_m)_i = y_i + v_i = a \cdot i + 1 \cdot b$$

Обработка многократно измеренных величин

- Составляем матрицу нормальных уравнений N и вектор c для системы нормальных уравнений $N \cdot k = c$

$$N = A^T \cdot A, \quad c = A^T \cdot y$$

- Решаем систему обращением матрицы

$$k = N^{-1} \cdot c = Q \cdot c$$

- Выявление значимости отличия от нуля a на основе t- критерия Стьюдента

$$t = \frac{a}{\sigma_a}, \quad t_{кр.} = t_{(1+p)/2,\ (n-2)} \quad t < t_{кр.}$$

Если неравенство выполняется – ряд достаточно случаен.

Обработка многократно измеренных величин

Случайность – непараметрический подход – мало измерений, неизвестен закон распределения измерений

Лучший критерий – критерий инверсий. Инверсия ряда

ki - число элементов, которые меньше предыдущего элемента. Коэффициент критерия – сумма инверсий *I*

Суть критерия – если все последующие элементы ряда меньше предыдущего, максимум инверсий – ряд полностью убывающий и наоборот (0 инверсий). В середине - ряд случаен – отличие числа инверсий от среднего значения. Нормировка

$$|I^*| = \frac{|I - M(I)|}{\sqrt{D(I)}} \le u_{\frac{1+p}{2}}$$

где $M(I)$ и $D(I)$ – известные мат. ожидание и дисперсия I

Обработка многократно измеренных величин

Проверка однородности результатов измерений.

Общий случай - сравнение законов распределений между собой - трудоёмко и часто невозможно. Сводят к сравнению главных характеристик распределения: сдвига, масштаба и наличия грубых измерений.

В геодезии для практики проверка однородности:

– проверить ряд на наличие грубых измерений;

– проверить на степень однородности дисперсии частей ряда измерений (проверка на гетероскедастичность);

– проверить на степень однородности сдвига некоторых частей ряда измерений (проверка на сдвиг центра).

Обработка многократно измеренных величин

Параметрические и непараметрические методы. Основой проверки на грубые измерения нормальная метка (z-метка) вида

$$z(x) = \frac{x - MO(x)}{\sqrt{D(x)}}$$

Параметрический критерий Смирнова-Греббса.

Статистика

$$z = \frac{\left| x_{экст.} - \overline{} \right|}{\sigma} \qquad \left\{ z_{\max.} = \frac{x_{\max.} - \bar{x}}{\sigma}; \quad z_{\min.} = \frac{\bar{x} - x_{\min.}}{\sigma} \right\}$$

Если $z > z_{\alpha,n}$ то с $P = 1 - \alpha$ выброс.

Не устойчив. Лучшие оценки параметров

Обработка многократно измеренных величин

Критерий ***Романовского*** – вычисление характеристик без подозреваемых:

$$\frac{\overline{x}' - x_{(1)}}{m'} > t_q \qquad \frac{x_{(n)} - \overline{x}'}{m'} > t_q$$

Неравенство выполняется – крайние грубые.

Критерий Ирвина по соседним в вариационном ряду:

$$\lambda = \frac{x_{(n)} - x_{(n-1)}}{m} \qquad \lambda = \frac{x_{(2)} - x_{(1)}}{m}$$

Если $\lambda > \lambda_P$ - крайние грубые.

Обработка многократно измеренных величин

Непараметрические методы оценивания грубых погрешностей – устойчивые оценки сдвига и масштаба – робастные метки.

Основной - критерий Хоглина-Иглевича (у нас правило Хэмпэла).

В нормальной метке – сдвиг-медиана, масштаб - абсолютное медианное отклонение (АМО, англ. MAD)

$$z_R(x) = \frac{|\, x - med(x)\,|}{med(|\, x - med(x))\,|} = \frac{|\, x - med(x)\,|}{(\quad)}$$

Чтобы определить ЗР метки - переводят $AMO(x)$ в стандартное отклонение теоретическим коэффициентом

Обработка многократно измеренных величин

$$c = \frac{1}{F^{-1}(0.75)} = 1.4826$$

$F^{-1}(p)$ - квантиль нормального закона распределения для вероятности p. Взяв вероятность 0.999 - предельный коэффициент 3.5

$$|\, x - med(x) \,| \le 3.5 \cdot 1.4826 \cdot AMO(x) = 5.2 \cdot AMO(x)$$

Устойчив, эффективен. Другой вид через границы

$$\left[med(x) - 5.2 \cdot AMO(x); \quad med(x) + 5.2 \cdot AMO(x) \right]$$

Все что выходит за границы – грубое. Другие параметризации характеристик сдвига и масштаба.

Обработка многократно измеренных величин

Проверка ряда измерений на однородность по главным характеристикам.

Делят на 2 или более части, используют параметрические и непараметрические критерии.

В параметрических критериях предполагается НЗР. Тогда

- для однородности масштаба используют обычный критерий отношений дисперсий (квадратов стандартного отклонения ,или F-критерий Фишера,

- для однородности сдвигов (степень отличия центров распределения) t-критерий Стьюдента.

Обработка многократно измеренных величин

Неравноточность дисперсий по критерию Фишера разбиением выборки на 2 подвыборки.

Статистика

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

Если $F < F_{эт}(p, f_1, f_2)$ ряд равноточен с вероятностью p.

Неоднородность средних по критерию Стьюдента с тем же разбиением. Статистика

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{ср}}$$

Если $t < t_{эт}((p+1)/2, f_1)$ ряд однороден по сдвигу (положению) с вероятностью p

Обработка многократно измеренных величин

Непараметрические критерии - закон распределения не известен, мало измерений.

Основной - критерий ранговой корреляции Спирмена (гетероскедастичность, неравноточность) элементов ряда.

Суть критерия – оценка меры рассеивания результатов измерений в виде остатков от модели (среднего). Есть выраженная неравноточность – есть постоянное увеличение (уменьшение) остатков с увеличением номера измерения i. Степень связанности номера и остатка - коэффициент корреляции. Нет неравенства дисперсий измерений – нет корреляции.

Обработка многократно измеренных величин

Для устойчивости - ранговый коэффициент корреляции Спирмена.

Реализация критерия:

- получаем оценки рассеивания измерений в виде остатков v_i

$$v_i = x_i - \overline{x}$$

-находим ранги ряда остатков

Ранг элемента – его номер в вариационном ряде.

- находим разность рангов $d_i = i - n_i$ и вычисляем коэффициент ранговой корреляции Спирмена

$$r_{i,v} = 1 - 6 \cdot \frac{\left[d^2\right]}{n(n^2 - 1)}$$

.

Обработка многократно измеренных величин

Окончание критерия - оценка отличия от нуля коэффициента корреляции между номером элемента в ряде и остатком. Используют t-критерий Стьюдента:

практика

$$t = \frac{r_{i,v}\sqrt{n-2}}{\sqrt{1-r_{i,v}^2}}$$

теория (эталон)

$$t_{эт}((1+p)/2,\ n-2).$$

Если $t < t_{эт}$ нулевая гипотеза об отсутствии гетероскедастичности (неравенства дисперсий измерений в ряде) не отвергается с вероятностью p.

Обработка многократно измеренных величин

Исследование на независимость элементов в ряде измерений.

Предполагает отсутствие значимой связи между элементами ряда с каким либо сдвигом (лагом). Определяется автокорреляцией лага 1.

Самый известный и используемый тест - критерий Дарбина-Уотсона.

Суть – установить значимость тесноты связи между рядом стоящими элементами ряда измерений.

Для вычисления критерия Дарбина-Уотсона выполняют следующие шаги:

Обработка многократно измеренных величин

– используя любой способ строят линейную модель для ряда измерений, чаще вида $(y_i)_{мод} = a \cdot i + b$ и находят величины остатков $v_i = (y_i)_{мод} - y_i$;

– по величинам остатков v_i вычисляют статистику DW критерия *Дарбина-Уотсона* как характеристику тесноты связи между рядом стоящими элементами ряда измерений y_i

$$DW = \frac{\sum_{i=2}^{n}(v_i - v_{i-1})^2}{\sum_{i=1}^{n} v_i^2} \approx 2(1 - r_{i,i-1})$$

откуда для коэффициента корреляции имеем

$$r_{v_i, v_{i-1}} \approx 1 - \frac{DW}{2}$$

Обработка многократно измеренных величин

Анализ результатов тестирования (см. формулы выше):

$$0 \le DW \le 4$$

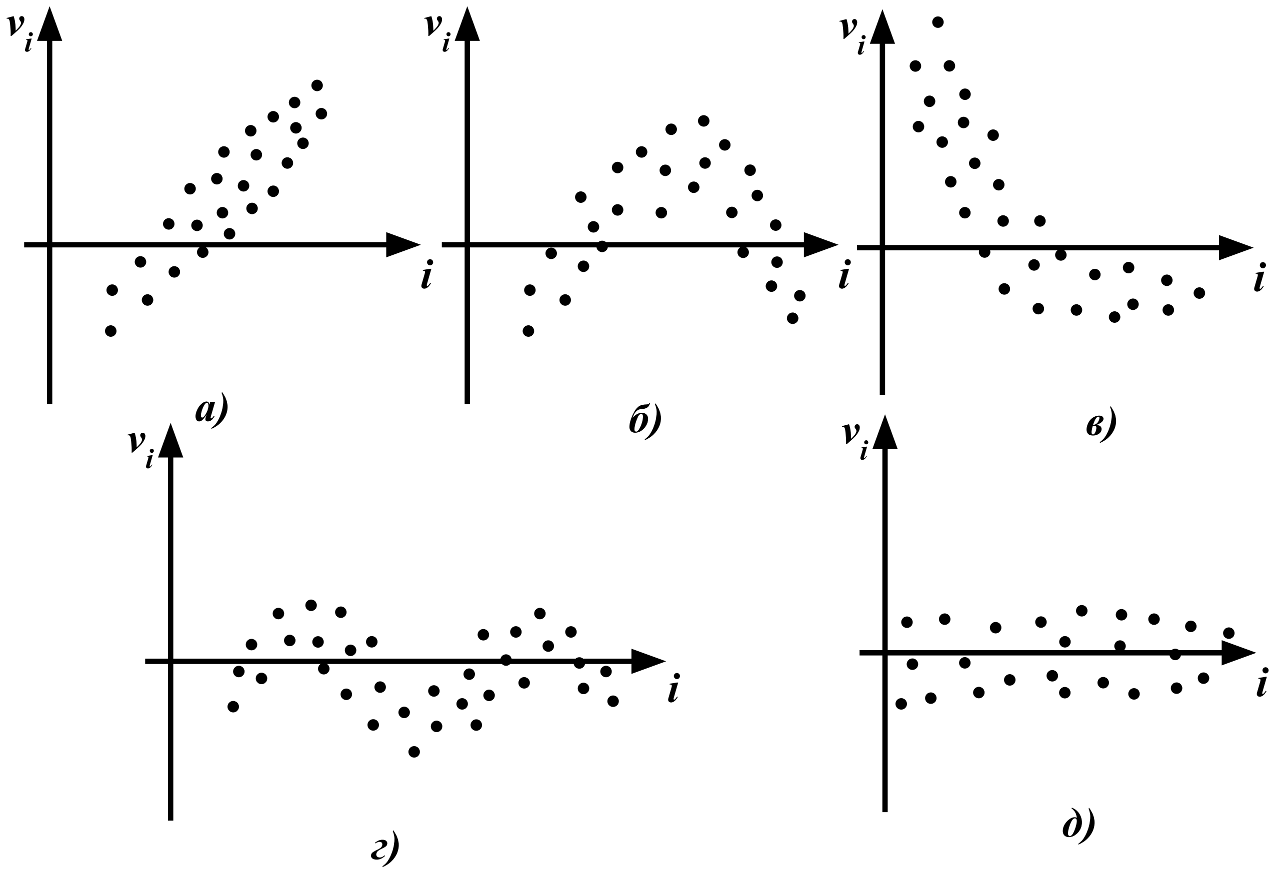
если $r_{v_i, v_{i-1}} \approx 0$ (отсутствие автокорреляции), то $DW \approx 2$,
если $r_{v_i, v_{i-1}} \approx 1$ (положительная автокорреляция), то $DW \approx 0$,
$r_{v_i, v_{i-1}}$
если $r_{v_i, v_{i-1}} \approx -1$ (отрицательная автокорреляция), то $DW \approx 4$.
Есть таблицы. Вычисления сложны.

Можно считать (но грубо), что если $1.5 < DW < 2.5$, то автокорреляция отсутствует.

Обработка многократно измеренных величин

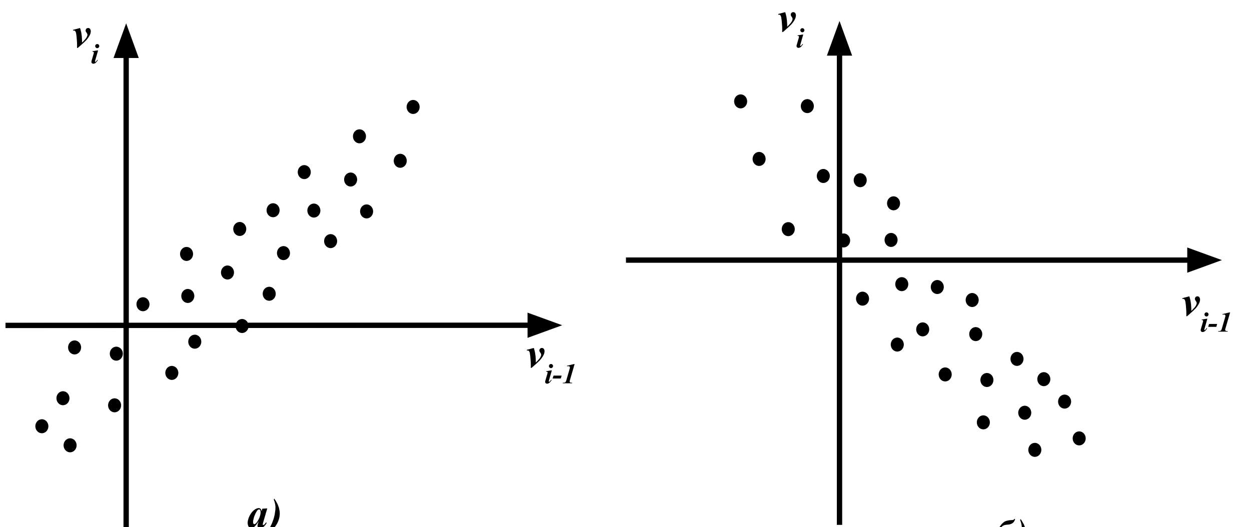
Некоторые графические возможности анализа:

Автокорреляция через последовательно-временные графики:



Обработка многократно измеренных величин

Положительная – отрицательная автокорреляция



а) *б)*

Обработка многократно измеренных величин

Выявление гетероскедастичности: