Федеральное агентство по образованию

Томский государственный
архитектурно-строительный университет

ИСПОЛЬЗОВАНИЕ МЕТОДА ИСКЛЮЧЕНИЯ ГРУБЫХ ОШИБОК В СТАТИСТИКЕ С ПРИМЕНЕНИЕМ ПЭВМ

Методические указания к лабораторным занятиям

Составитель М.В. Анисимов

Томск – 2008

Использование метода исключения грубых ошибок в статистике с применением ПЭВМ: методические указания / Сост. М.В. Анисимов. – Томск.: Изд-во Том. гос. архит.-строит. ун-та, 2008. – 17 с.

Рецензент д.т.н., С.А. Карауш
Редактор Е.Ю. Глотова

Методические указания к лабораторным занятиям по дисциплинам ЕН.В.2 «Численные методы решения задач БЖД» для студентов специальности 280102 «Безопасность технологических процессов и производств» очной формы обучения.

Печатаются по решению методического семинара кафедры охраны труда и окружающей среды, протокол № 10 от 05.10.2008 г.

Утверждены и введены в действие проректором по учебной работе В.В. Дзюбо

с 01.11.2008
до 01.11.2013

Оригинал-макет подготовлен автором

Подписано в печать 01.11.2008.
Формат 60х90/16. Бумага офсет. Гарнитура Таймс,
печать офсет. Уч-изд. л. 0,9. Тираж 100 экз. Заказ №

Изд-во ТГАСУ, 634003, г. Томск, пл. Соляная, 2.
Отпечатано с оригинал-макета в ООП ТГАСУ.
634003, г. Томск, ул. Партизанская, 15.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Зачастую инженеру, в том числе инженеру в области обеспечения БЖД на производстве, необходимо либо получать эмпирические данные либо обрабатывать результаты проведенных экспериментов, а также статистические данные. Это требует от инженера умения не только грамотно пользоваться методами исключения различного вида погрешностей при проведении исследований, но и умения использовать возможности ПЭВМ для ускорения вычислительных процессов, а также увеличения объемов исследуемой информации.

Численное значение физической величины получается в результате ее измерения, т. е. сравнения ее с другой величиной того же рода, принятой за единицу. При выбранной системе единиц результаты измерений выражаются определенными числами. Известно, что при достаточно точных измерениях одной и той же величины результаты отдельных измерений отличаются друг от друга и, следовательно, содержат ошибки.

Ошибкой измерения называется разность $x - a$ между результатом измерения x и истинным значением a измеряемой величины. Ошибка измерения обычно неизвестна, как неизвестно и истинное значение измеряемой величины (исключения составляют измерения известных величин, проведенные со специальной целью исследования ошибок измерения, например для определения точности измерительных приборов). Одной из основных задач математической обработки результатов эксперимента как раз и является оценка истинного значения измеряемой величины по получаемым результатам. Другими словами, после неоднократного измерения величины a и получения ряда результатов, каждый из которых содержит некоторую неизвестную ошибку, ставится задача вычисления приближенного значения a с возможно меньшей ошибкой. Для решения этой задачи (при данном уровне точности измерений) надо знать основные свойства ошибок измерений и уметь ими воспользоваться.

1. ОБЩИЕ ПОЛОЖЕНИЯ

Грубые ошибки. Прежде всего, при математической обработке результатов измерений не следует учитывать заведомо неверные результаты (промахи), или, как говорят, результаты, содержащие грубые ошибки. Грубые ошибки возникают вследствие нарушения основных условий измерения или в результате недосмотра экспериментатора (например, при плохом освещении вместо «3» записывают «8»). При обнаружении грубой ошибки результат измерения следует сразу отбросить, а само измерение повторить (если это возможно). Внешним признаком результата, содержащего грубую ошибку, является его резкое отличие по величине от результатов остальных измерений. На этом основаны некоторые критерии исключения грубых ошибок по их величине, однако самым надежным и эффективным способом браковки неверных результатов остается браковка их непосредственно в процессе самих измерений.

Систематические ошибки. Ошибки измерения вызываются большим количеством разнообразных причин (факторов). Иногда в проведенной серии измерений удается выделить такие причины ошибок, эффект действия которых может быть рассчитан. Например, если после измерений обнаружена неправильная регулировка прибора, которая привела к смещению начала отсчета, то все снятые показания будут смещены либо на постоянную величину, если шкала прибора равномерна, либо на величину, изменяющуюся по определенному закону, если шкала прибора неравномерна. Другим примером может служить изменение внешних условий, например, температуры, если известно влияние этих изменений на результаты измерений. К названным причинам можно также отнести некоторое несовершенство измерительных приборов на границе области их применимости, вызывающее известные ошибки.

Принято говорить, что каждая из таких причин вызывает систематическую ошибку. Выявление систематических ошибок, вызываемых каждым отдельным фактором, требует специальных исследований (например, измерений одной и той же величины разными методами или измерение одним и тем же прибором некоторых эталонов, известных величин). Но как только систематические ошибки обнаружены, и их величины рассчитаны, они могут быть легко устранены путем введения соответствующих поправок в результаты измерения. Подчеркнем, что при этом общая ошибка каждого результата остается неизвестной, так что речь идет не о выделении из общей ошибки некоторой части в виде систематической ошибки, а лишь о введении поправок на известный эффект действия тех факторов, которые удалось выявить.

Случайные ошибки. Ошибки измерения, остающиеся после устранения всех выявленных систематических ошибок, т. е. ошибки результатов измерений, исправленные путем введения соответствующих поправок, называются случайными. Случайные ошибки вызываются большим количеством таких факторов, эффекты действия которых столь незначительны, что их нельзя выделить и учесть в отдельности (при данном уровне техники и точности измерений). Случайную ошибку можно рассматривать как суммарный эффект действия таких факторов. Случайные ошибки являются неустранимыми, их нельзя исключить в каждом из результатов измерений. Но с помощью методов теории вероятностей можно учесть их влияние на оценку истинного значения измеряемой величины, что позволяет определить значение измеряемой величины со значительно меньшей ошибкой, чем ошибки отдельных измерений. Учет влияния случайных ошибок основан на знании законов их распределения.

Показатели точности измерения

Параметр σ («сигма») называется средней квадратической ошибкой измерения, стандартной ошибкой или просто стандартом. Квадрат величины σ называется дисперсией ошибки.

Кроме названных, иногда применяются и другие показатели точности измерений. Соотношения между различными показателями точности в случае нормального распределения ошибок следующие.

Вероятная ошибка

$$\rho = 0{,}6745\sigma, \qquad 2\Phi(\rho) = 0{,}5 \tag{1}$$

Средняя абсолютная ошибка

$$\vartheta = \int_{-\infty}^{\infty} |z| \rho(z) dz = \sigma \frac{2}{\sqrt{2\pi}} = 0{,}7979\sigma. \tag{2}$$

Мера точности

$$h = \frac{1}{\sigma\sqrt{2}} = 0{,}7071 \frac{1}{\sigma}. \tag{3}$$

2. МЕТОДЫ ИСКЛЮЧЕНИЯ ГРУБЫХ ОШИБОК

При получении результата измерения, резко отличающегося от всех других результатов, естественно, возникает подозрение, что допущена грубая ошибка. В этом случае необходимо сразу же проверить, не нарушены ли основные условия измерения.

Если же такая проверка не была сделана вовремя, то вопрос о целесообразности браковки одного «выскакивающего» значения решается путем сравнения его с остальными результатами измерения. При этом применяются различные критерии, в зависимости от того, известна или нет средняя квадратическая ошибка *a* измерений (предполагается, что все измерения производятся с одной и той же точностью и независимо друг от друга).

2.1. Метод исключения при известной σ

Обозначим «выскакивающее» значение через x_*, а все остальные результаты измерения через x_1, x_2, ... , x_n. При этом среднее арифметическое значение будет определяться как

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}. \qquad (4)$$

Сравним абсолютную величину разности $x_* - \overline{\mathrm{x}}$ с величиной $\sigma\sqrt{(n+1)/n}$.

Для полученного отношения

$$t = \frac{\left|x_* - \bar{x}\right|}{\sigma\sqrt{(n+1)/n}} \qquad (5)$$

подсчитываем вероятность 1 – 2Ф(t) с помощью табл. 1 приложения 1.

Это позволяет оценить вероятность того, что рассматриваемое отношение случайно примет значение, не меньшее чем *t*, при условии, что значение x_* не содержит грубой ошибки (что ошибка результата x_* только случайна). Если подсчитанная указанным образом вероятность оказывается очень малой, то «выскакивающее» значение содержит грубую ошибку и его следует исключить из дальнейшей обработки результатов измерений.

Какую именно вероятность считать очень малой, зависит от конкретных условий решаемой задачи: если назначить слишком низкий уровень малых вероятностей, то грубые ошибки могут остаться, если же взять этот уровень неоправданно большим, то можно исключить результаты со случайными ошибками, необходимые для правильной обработки результатов измерения. Обычно применяют один из трех уровней малых вероятностей: 5 %-й уровень (исключаются ошибки, вероятность появления которых меньше 0,05); 1 %-й уровень (исключаются ошибки, вероятность появления которых меньше 0,01); 0,1 %-й уровень (исключаются ошибки, вероятность появления которых меньше 0,001). При выбранном уровне a малых вероятностей «выскакивающее» значение x_* считают содержащим грубую ошибку, если для соответствующего отношения t (5) вероятность $1 - 2\Phi(t) < \alpha$. Чтобы подчеркнуть вероятностный характер этого заключения, говорят, что значение x_* содержит грубую ошибку с надежностью вывода $P = 1 - \alpha$. Значение $t = t(P)$, для которого $1 - 2\Phi(t) = \alpha$, и значит, $2\Phi(t) = P$, называется критическим значением отношения (5) при надежности P. Так, если $\alpha = 0{,}01$ (1 %-й уровень), то $P = 0{,}99$, критическое значение $t = t(P) = 2{,}576$ (см. табл. 1, прил.), и как только отношение (5) превзойдет это критическое значение, мы можем браковать «выскакивающее» значение x_* с надежностью вывода 0,99.

2.2. Метод исключения при неизвестной σ

Если величина σ заранее неизвестна, то она оценивается приближенно по результатам измерений, т. е. вместо нее применяют эмпирический стандарт

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{6}$$

При этом абсолютную величину разности $|x_* - \bar{x}|$ между «выскакивающим» значением x_* и средним значением $\bar{x}$ остальных (приемлемых) результатов делят на эмпирический стандарт и полученное отношение

$$t = \left|x_* - \bar{x}\right| / s \tag{7}$$

сравнивают с критическими значениями t_n (P) из табл. 2 (приложения). Если при данном числе n приемлемых результатов отношение (7) оказывается между двумя критическими значениями при надежностях P_1 и P_2 ($P_2 > P_1$), то с надежностью вывода, большей P_1, можно считать, что «выскакивающее» значение содержит грубую ошибку, и исключить его из дальнейшей обработки результатов.

Заметим, что если надежность вывода окажется недостаточной, то это свидетельствует не об отсутствии грубой ошибки, а лишь об отсутствии достаточных оснований для исключения «выскакивающего» значения.

3. СРЕДНИЕ ЗНАЧЕНИЯ ВЫБОРКИ И ИХ ОЦЕНКИ

Средним арифметическим значением (или просто средним значением) величин $x_1, x_2, \ldots x_n$ называется

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{8}$$

Средним квадратическим отклонением величин $x_1, x_2, \ldots x_n$ от их среднего значения $\bar{x}$ называется

$$s^* = \sqrt{\frac{(x_1 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{9}$$

При вычислении среднего квадрата отклонения от любого числа c имеет место зависимость

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + (\bar{x} - c)^2. \tag{10}$$

Если среди результатов измерений встречаются равные, то соответствующие слагаемые в формулах (8) и (9) можно объединить.

Пусть результат x_1 встречается m_1 раз, результат x_2 встречается m_2 раз, ..., результат $x_к$ встречается $m_к$ раз ($m_1 + m_2 + ... + m_к = n$).

Тогда зависимости (8) и (9) примут вид

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} m_i x_i,$$

$$s^* = \sqrt{\frac{1}{n}\sum_{i=1}^{k} m_i (x_i - \bar{x})^2}. \qquad (11)$$

4. ПРИМЕР ПРИМЕНЕНИЯ МЕТОДА ИСКЛЮЧЕНИЯ ГРУБЫХ ОШИБОК ПРИ РЕШЕНИИ ЗАДАЧ БЖД

Пример решения методом исключения при известной σ.

Пусть среди 10 результатов независимых измерений, произведенных со средней квадратической ошибкой σ = 0,133, обнаружено одно «выскакивающее» значение $x_* = 4{,}5$, в то время как среднее из остальных 9 результатов составляет $\bar{x}$ = 3,69 (среднее значение вычислено по зависимости (8)).

№	1	2	3	4	5	6	7	8	9	10
x	3,75	3,65	4,5	3,75	3,70	3,65	3,70	3,65	3,60	3,80

Можно ли считать, что «выскакивающее» значение содержит грубую ошибку, и исключить его из дальнейшей обработки?

Решение. Разность между «выскакивающим» значением и средним составляет $x_* - \bar{x} = 0{,}81$, поэтому отношение (5) равно

$$t = \frac{\left| x_* - \bar{x} \right|}{\sigma\sqrt{(n+1)/n}} = \frac{\left|4{,}5 - 3{,}69\right|}{0{,}133\sqrt{10/9}} = 5{,}7.$$

По табл. 1 для t = 5,0 оцениваем вероятность

$$1 - 2\Phi\,(t) = 0{,}0000006 < 0{,}001.$$

Следовательно, с надежностью вывода P > 0,9999 (при уровне малой вероятности α = 0,1%) можно считать, что значение x_* содержит грубую ошибку, и исключить это значение из дальнейшей обработки результатов измерения.

Пример решения методом исключения при неизвестной σ.

Пусть для данных задачи 1 эмпирический стандарт будет равен s = 0,133. Можно ли исключить результат x_* = 4,5 из дальнейшей обработки?

Решение. Здесь отношение (7) равно

$$t = |0{,}81| \,/\, 0{,}133 = 6{,}1.$$

Если число приемлемых результатов n = 10, то полученное отношение превосходит критическое значение 6,1 при надежности $P \geq 0{,}999$ и значение x_* = 4,5 можно исключить с надежностью вывода, большей 0,999 (см. табл. 2, прил.).

5. ЗАДАНИЕ ДЛЯ ЛАБОРАТОРНОЙ РАБОТЫ

На основании данных, представленных в таблицах ниже, выполнить анализ числовой выборки методом исключения грубых ошибок в следующей последовательности:

1. Решить поставленную задачу численно в тетради.
2. С помощью Excel выполнить устранение погрешности предложенной числовой выборки методом исключения грубых ошибок.
3. С помощью Excel выполнить расчет среднеарифметического значения и среднеквадратичного отклонения значений в

предложенной числовой выборке.

4. В графическом виде представить результаты проделанной работы. График выполнить в приложении Origin 5.0.

Данные для расчетов

Вариант 1

№	1	2	3	4	5	6	7	8	9	10
y	1.860	1.87	1.84	1.89	1.87	2.30	1.845	1.850	1.865	1.867

Вариант 2

№	1	2	3	4	5	6	7	8	9	10
y	2,6	2,65	2,55	2,6	3,80	2,7	2,66	2,65	2,45	2,55

Вариант 3

№	1	2	3	4	5	6	7	8	9	10
y	3,75	3,65	3,7	4,35	3,80	3,78	3,8	3,75	3,7	3,75

Вариант 4

№	1	2	3	4	5	6	7	8	9	10
y	2,0	2,2	2,1	1,8	2,1	2,0	1,9	3,1	2,1	1,9

Вариант 5

№	1	2	3	4	5	6	7	8	9	10
y	3,5	3,45	3,55	4,1	3,55	3,5	3,47	3,52	3,48	3,49

Вариант 6

№	1	2	3	4	5	6	7	8	9	10
y	4,0	3,95	3,98	4,05	5,0	4,1	4,05	3,95	4,0	4,10

Вариант 7

№	1	2	3	4	5	6	7	8	9	10
y	3,0	3,1	3,1	3,0	4,4	2,9	2,96	2,97	2,98	2,9

Вариант 8

№	1	2	3	4	5	6	7	8	9	10
y	4,0	3,9	3,95	3,98	4,1	4,1	5,1	3,95	4,,05	4,05

Вариант 9

№	1	2	3	4	5	6	7	8	9	10
y	24	24,1	24,2	23,9	27	24,9	24,1	24,2	24,05	23,9

Вариант 10

№	1	2	3	4	5	6	7	8	9	10
y	15,5	15,4	15,5	17,0	15,3	15,4	15,5	15,6	15,65	15,3

Вариант 11

№	1	2	3	4	5	6	7	8	9	10
y	3,75	3,65	3,7	4,35	3,80	3,78	3,8	3,75	3,7	3,75

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Перечислите виды статистических ошибок.
2. Какие три уровня малых вероятностей используется при проведении исключения погрешностей измерения?
3. Охарактеризуйте формулу определения среднего арифметического значения.
4. Охарактеризуйте формулу расчета среднеквадратичного отклонения.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Вержбицкий, В.Н. Основы численных методов / В.Н. Вержбицкий. – М. : Высшая школа, 2005. – 840 с.

2. Кулеш, У. Достоверные вычисления. Базовые численные методы / У. Кулеш. – РХД, 2005. – 4 с.

3. Самарский, А.А. Численные методы математической физики / А.А. Самарский, А.В. Гулин. – М. : Научный мир, 2000. – 315 с.

4. Репин, С.В. Математические методы обработки статистической информации с помощью ЭВМ / С.В. Репин. – Минск: Университетское, 1990. – 127 с.

5. Турчак, Л.И. Основы численных методов: учебное пособие для вузов / Л.И. Турчак, П.В. Плотников. – М. : Физматлит, 2002. – 304 с.

6. Гришин, В.К. Математическая обработка и интерпретация физического эксперимента / В.К. Гришин, Ф.А. Живописцев, В.А. Иванов. – М. : Изд-во МГУ, 1988. – 317 с.

7. Браунс, Я.А. Обработка результатов и планирование эксперимента: учебное пособие по УИРС для студентов / Я.А. Браунс. – Рига: Рижский политехн. ин-т, 1989. – 68 с.

8. Введение в математическое моделирование: учебное пособие / В.Н. Ашихмин, М.Г. Бояршинов, М.Б. Гитман, И.Э. Келлер [и др.]. – М. : «Интернет Инжиниринг», 2000. – 336 с.

Приложение

Таблица 1

t	$\Phi(t)$	$1 - 2\Phi(t)$	$1 - P$	$t = t\,(P)$	P
2,5	0,49379	0,01242	0,05	1,960	0,95
2,6	0,49534	0,00932	0,04	2,054	0,96
2,7	0,49653	0,00693	0,03	2,170	0,97
2,8	0,49744	0,00511	0,02	2,326	0,98
2,9	0,49813	0,00373	0,01	2,576	0,99
3,0	0,49865	0,00270	0,009	2,612	0,991
3,1	0,49903	0,00194	0,008.	2,652	0,992
3,2	0,49931	0,00137	0,007	2,697	0,993
3,3	0,49952	0,00097	0,006	2,748	0,994
3,4	0,49966	0,00067	0,005	2,807	0,995
3,5	0,499767	0,000465	0,004	2,878	0,996
3,6	0,499841	0,000318	0,003	2,968	0,997
3,7	0,499892	0,000216	0,002	3,090	0,998
3,8	0,499927	0,000145	0,001	3,291	0,999
3,9	0,499952	0,000096	0,0009	3,320	0,9991
4,0	0,499968	0,000063	0,0008	3,353	0,9992
4,1	0,499979	0,000041	0,0007	3,390	0,9993
4,2	0,499987	0,000027	0,0006	3,432	0,9994
4,3	0,499991	0,000017	0,0005	3,481	0,9995
4,4	0,499995	0,000011	0,0004	3,540	0,9996
4,5	0,4999966	0,0000068	0,0003	3,615	0,9997
4,6	0,4999979	0,0000041	0,0002	3,720	0,9998
4,7	0,4999987	0,0000025	0,0001	3,891	0,9999
4,8	0,4999992	0,0000016	10^{-5}	4,417	$1 \cdot 10^{-5}$
4,9	0,4999995	0,0000009	10^{-6}	4,892	$1 \cdot 10^{-6}$
5,0	0,4999997	0,0000006	10^{-7}	5,327	$1 \cdot 10^{-7}$

Таблица 2

Критические значения t_n (P) отношения (7) для браковки «выскакивающих» значений x_*

P / n	0,95	0,98	0,99	0,999	P / n	0,95	0,98	0,99	0,999
5	3,04	4,11	5,04	9,43	20	2,145	2,602	2,932	3,979
6	2,78	3,64	4,36	7,41	25	2,105	2,541	2,852	3,819
7	2,62	3,36	3,96	6,37	30	2,079	2,503	2,802	3,719
8	2,51	3,18	3,71	5,73	35	2,061	2,476	2,768	3,652
9	2,43	3,05	3,54	5,31	40	2,048	2,456	2,742	3,602
10	2,37	2,96	3,41	5,01	45	2,038	2,441	2,722	3,565
11	2,33	2,89	3,31	4,79	50	2,030	2,429	2,707	3,532
12	2,29	2,83	3,23	4,62	60	2,018	2,411	2,683	3,492
13	2,26	2,78	3,17	4,48	70	2,009	2,399	2,667	3,462
14	2,24	2,74	3,12	4,37	80	2,003	2,389	2,655	3,439
15	2,22	2,71	3,08	4,28	90	1,998	2,382	2,646	3,423
16	2,20	2,68	3,04	4,20	100	1,994	2,377	2,639	3,409
17	2,18	2,66	3,01	4,13	∞	1,960	2,326	2,576	3,291
18	2.17	2.64	2.98	4.07					