МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Юго-Западный государственный университет» (ЮЗГУ)

Кафедра фундаментальной химии и химической технологии

УТВЕРЖДАЮ
Проректор по учебной работе
_____________ О.Г. Локтионова
«_7_» _02_ __________ 2018 г.

ПРАКТИЧЕСКИЕ АСПЕКТЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ В ХИМИЧЕСКОЙ ПРАКТИКЕ

Методические указания к практической и самостоятельной работе по дисциплине «Статистическая обработка в химической практике» для студентов направлений 18.03.01 - Химическая технология

Курск 2018

УДК 66(076.5)
Составители: С.Д. Пожидаева, А.М. Иванов

Рецензент
Кандидат химических наук, доцент *Н.А. Борщ*

Практические аспекты статистической обработки в химической практике: методические указания к практической и самостоятельной работе по дисциплине «Статистическая обработка в химической практике» для студентов направлений 18.03.01 - Химическая технология / Юго-Зап.гос.ун-т; сост.: С.Д. Пожидаева, А.М. Иванов. Курск, 2018. 27 с. 14 табл.

В методические указания включены работы по статистической обработке экспериментальных данных в химическом эксперименте, позволяющие освоить методы и приемы оценки погрешностей измерений, испытаний и вычислений, основы корреляционного анализа и метода наименьших квадратов.

Методические указания к практическим работам по дисциплине «Статистическая обработка в химической практике» для студентов направлений 18.03.01 - Химическая технология.

Методические указания соответствуют требованиям программы.

Текст печатается в авторской редакции

Подписано в печать 7.02.18. Формат 64x18 1/16
Усл.печ.л.1,4 Уч.-изд.л.1,3 Тираж 100 экз. Заказ 614. Бесплатно
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

СОДЕРЖАНИЕ

Введение

Данные методические указания предназначены для создания необходимой теоретической и практической базы для восприятия и усвоения современных знаний в области теоретических основ химической технологии (ПАХТ, ОХТ, НИРС) и их практических приложений (спецкурсы), вычленения роли и места научных исследований в подготовке молодых специалистов. Полученные навыки должны студенту правильно обосновать выбор метода измерения или анализа, удовлетворяющий требованиям точности, быстроты и экономичности выполнения анализа с учетом возможной величины погрешности измерения, а также показать эффективность использования методов высшей математики в решении практических вопросов.

Используемые в методических указаниях подходы действенны для формирования действительного моста между фундаментальными и технологическими дисциплинами в плане преемственности, подходов и использования достижений отдельных дисциплин при решении практических задач.

В методических указаниях приводятся практические работы для закрепления теоретических знаний по дисциплине «Статистическая обработка в химической практике».

Практическая работа №1

СРЕДНИЕ ВЕЛИЧИНЫ И ИХ ИСПОЛЬЗОВАНИЕ В ХИМИЧЕСКОЙ ПРАКТИКЕ

Случайную погрешность можно оценить из серии параллельных (повторных) измерений. Например, если $X_1, X_2..X_n$— значения объемов стандартного раствора в серии *n независимых* повторных титрований, то эти значения обычно отличаются друг от друга вследствие погрешностей, неизбежно возникающих в ходе любой измерительной процедуры.

При этом $\overline{X}$ в n раз более точен по сравнению с X_I. Так как истинное значение определяемого содержания обычно неизвестно, то результат анализа $(\overline{X})$ сравнивают с действительным значением этой величины. За действительное значение измеряемой величины обычно принимают рассчитанное содержание определяемого компонента (в случае анализа химически чистого вещества), его содержание в стандартном образце или результат определения, полученный при помощи стандартного метода анализа.

При проведении измерения обычно не ограничиваются единичным измерением, а проводят несколько параллельных измерений (3-5) для одной и той же пробы в одинаковых условиях. Средний результат называют результатом анализа и обозначают $\tilde{O}$.

Если величина *X* определяет некоторое свойство совокупности, то средней величиной будет такое значение $\tilde{O}$, при замене на которое отдельных значений *X* это свойство совокупности не изменяется.

Средние значения величин могут определяться различными способами, выбор которых обусловлен связью между усредненными величинами и тем свойством, которое они определяют. Наиболее широко применяются на практике: 1) средняя арифметическая взвешенная; 2) простая средняя арифметическая; 3) медиана; 4) мода; 5) средняя логарифмическая; 6) средняя квадратичная; 7) средняя геометрическая; 8) средняя гармоническая.

Средняя арифметическая

Для совокупности величин $x_1, x_2, x_3, \ldots, x_t,$ которые в ней встречаются соответственно $n_1, n_2, n_3, \ldots, n_t$ раз. Тогда

$$\bar{X} = \frac{n_1x_1 + n_2x_2 + \ldots + n_tx_t}{n_1 + n_2 + \ldots + n_t} = \frac{\sum n_ix_i}{\sum n_i} \tag{1}$$

где $\bar{X}$ - это средняя арифметическая взвешенная; $n_1, n_2, \ldots, n_t$ – веса или частоты значений $x_1, x_2, \ldots, x_t$. Если все частоты значений равны 1 ($n_i=1$), то $\bar{X}$ является простой средней арифметической и рассчитывается как

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_t}{t} = \frac{\sum\limits_{i=1}^{t} x_i}{t} \tag{2}$$

Медиана

Во многих случаях *медиана* является более реалистичной характеристикой «центральной» величины, чем среднее. Медиана – такое среднее значение, которое делит совокупность значений x_i на две равные по количеству членов части, причем в одной из них все значения x_i меньше медианы, а в другой – больше.

Если расположить все члены совокупности в ряд в возрастающем порядке, то при нечетном числе членов, т.е. при n=2m+1, медианой будет значение среднего члена ряда, т.е. Me = x_{m+1}. При четном числе членов ряда, т.е. n=2m, за медиану принимается среднее арифметическое двух значений x_m и x_{m+1}, находящихся в середине ряда, т.е.

$$Me = \left(x_m + x_{m+1}\right)/2 \qquad (3)$$

Если в данном ряду члены, достаточно удаленные от медианы, подвергаются немалым изменениям, то медиана при этом не меняется, в то время как средняя арифметическая изменяется. Поэтому, если, как это часто бывает, значения x_i , находящиеся на концах ряда, не точны, то в качестве средней лучше пользоваться медианой, а не средней арифметической.

Мода

Модой называется наиболее вероятное значение случайной величины или то значение случайной величины, частота которого наибольшая. Мода используется для характеристики наиболее часто встречающихся значений в совокупности случайных величин. Например, при размоле руды мода характеризует размер зерен фракции, выход которой наибольший, что служит критерием работы размольной машины.

Средняя логарифмическая

В химической практике одними из наиболее распространенных являются логарифмические зависимости (кинетика реакций первого порядка, pH-метрия и т.д.). В этих случаях кривая распределения имеет логарифмический характер и величиной, характеризующей среднее значение, является средняя логарифмическая.

Средняя логарифмическая двух величин есть отношение их разности к разности их натуральных логарифмов

$$X_{\text{ñð.ëîã.}} = \frac{x_1 - x_2}{\ln x_1 - \ln x_2} = \frac{x_1 - x_2}{\ln\left(x_1/x_2\right)} \qquad (4)$$

Средняя квадратичная

Средней квадратичной n положительных или отрицательных величин $x_1, x_2, \dots , x_n$ называется положительное значение квадратного корня из суммы квадратов этих величин, деленной на их число

$$X_{\text{ср.êâ.}} = \sqrt{\left(x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2\right)/n} \qquad (5)$$

Средняя геометрическая

Средней геометрической n положительных величин $x_1, x_2, \dots ,$ x_n называется положительное значение корня n-й степени из их произведения

$$X_{\text{ñð.ãåîì.}} = +\sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \qquad (6)$$

Среднюю геометрическую удобнее вычислять из формулы, которая получается логарифмированием формулы (6)

$$\lg X_{ср.геом.} = (\lg x_1 + \lg x_2 + \ldots + \lg x_n)/n \qquad (7)$$

Средняя гармоническая

Средней гармонической n положительных величин $x_1, x_2, \ldots, x_n$, называется величина H, обратное значение которой равно среднему арифметическому обратных значений величин $x_1, x_2, \ldots, x_n$, т.е.

$$(1/H) = [(1/x_1) + (1/x_2) + \ldots + (1/x_n)]/n. \qquad (8)$$

или $$H = \frac{n}{(1/x_1) + (1/x_2) + \ldots + (1/x_n)}. \qquad (9)$$

Средняя гармоническая используется в тех случаях, когда приходится иметь дело с величиной, зависящей от обратных значений частных величин.

Примеры заданий с решениями.

1. При определении содержания Na_2CO_3 в растворе соды (100 мл) путем прямого титрования его аликвотных частей раствором HCl были получены следующие результаты (в г): 0,2031; 0,2033; 0,2015; 0,2048; 0,2020.

Найти среднее арифметическое.

Решение. Найдем среднее арифметическое результатов:

X = (0,2031 + 0,2033 + 0,2015 + 0,2048+0,2020) /5=0,2029 г.

2. Для полученных данных вычислить медиану

10,54;10,40; 10,10; 10,60; 10,50; 10,90; 10,20; 10,80; 10,46;

Решение. После расположения значений в порядке возрастания, находим, что медиана равна пятому по порядку значению, а именно, 10,50.

3. Для полученных данных вычислить медиану

10,54;10,40; 10,10; 10,60; 12,80, 10,50; 10,90; 10,20; 10,80; 10,46.

Решение. По аналогии с рассмотренным выше случаем, располагаем значения в порядке возрастания. Для полученного ряда значений медиана станет равна

(10,50 + 10,54)/2 = 10,52.

4. Найти медиану для выборки результатов, полученных при определении содержания Na_2CO_3 в растворе соды (100 мл) путем прямого титрования его аликвотных частей раствором HCl. Результаты (в г): 0,2031; 0,2033; 0,2015; 0,2048; 0,2020.

Решение. Располагают значения выборки в возрастающем порядке. Медианой выборки является значение 0,2031.

Задачи

1. При титровании 5 мл раствора щелочи 0,075 н. раствором со-

ляной кислоты, были получены следующие значения объемов. Определить среднее значение концентрации раствора щелочи.

V, мл	7,3	7,9	7,9	7,0	7,4	7,1	7,2	7,2	7,0	7,5	7,4	7,9	7,2	7,5	7,1	8,0
№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

2. При титровании раствора кислоты пипеткой объемом 25 мл с помощью раствора КОН с концентрацией 0,15 моль/л, получены следующие результаты: 15,3; 15,2; 15,7; 14,9; 14,9; 15,0; 15,2; 15,2; 15,3; 14,9; 14,9 мл. Определить среднее значение концентрации кислоты

3. Проведенные при неизменной температуре измерения парциального давления окислов азота над нитрозой, позволили получить значения: 3,52; 3,33; 3,60; 3,36; 3,14; 3,31; 3,34; 3,37; 3,39; 3,44; 3,35; 3,69; 3,12; 3,25.

Определить среднее арифметическое значение этих значений и медиану. Какое из этих значений, по вашему мнению, будет менее надежным и почему?

4. Ситовый анализ молотой руды дал следующие результаты

крупность зерен	550 – 500	500 – 450	450-400	400 – 350	350-300	300 – 250	250-200	200 – 150	150-100
выход,%	14,8	5,0	12,3	8,9	14,2	24,9	13,6	5,9	22,8

Определить величину моды.

Многовариантные задачи

1. Для полученного ряда значений (таблица 1) определить моду, медиану, среднее арифметическую, среднюю геометрическую, среднюю гармоническую. Обосновать выбор величины для характеристики среднего значения.

Таблица 1 – Задание для вычисления

1	0,51; 0,50; 0,51; 0,54; 0,52; 0,55; 0,50; 0,52; 0,50; 0,51; 0,55; 0,52; 0,52;
2	0,11; 0,11; 0,12; 0,14; 0,11; 0,12; 0,115; 0,13; 0,13;0,12; 0,11; 0,12; 0,13;
3	0,33; 0,33; 0,34; 0,34; 0,33; 0,34; 0,35; 0,33; 0,33;0,34; 0,33; 0,35; 0,36;
4	3,12; 3,14; 3,25; 3,31; 3,33; 3,34; 3,35; 3,36; 3,37; 3,39; 3,44; 3,52; 3,60;
5	7,5; 7,6; 7,4; 7,3; 7,2; 7,5; 7,2; 7,5; 7,6; 7,5; 7,3; 7,5; 7,6; 7,4; 7,8
6	17,45; 17,38; 17,65; 17,54; 17,12; 17,65; 17,45; 17,69; 17,12; 17,38; 17,89;
7	0,23;0,24; 0,28; 0,29; 0,23; 0,27; 0,32; 0,31; 0,32; 0,27;0,28; 0,35; 0,30;
8	4,54; 4,56; 4,87; 4;98; 4,12; 4,13; 4,45; 4,68; 4,63; 4,52; 4,96; 4,32; 4,68;
9	9,78; 9,77; 9,68; 9,64; 9,99; 9,87; 9,21; 9,99; 9,67; 9,86; 9,63; 9,45; 9,32;
10	1,56; 1,53; 1,56; 1,45; 1,12; 1,78; 1,98; 1,35; 1,64; 1,68; 1,45; 1,34; 1,39;
11	0,10; 0,11; 0,12; 0,12; 0,11; 0,10; 0,095; 0,095; 0,10; 0,090; 0,089; 0,087;
12	4,54; 4,58; 4,32; 4,36; 4,87; 4,65; 4,454 4,63; 4,14; 4,32; 4,35; 4,19; 4,68;

2. Для полученного ряда значений определить концентрацию

растворов, для которых определить моду, медиану, среднее арифметическую, среднюю геометрическую, среднюю гармоническую.

Ряд 1	Ряд 2	Ряд 3	Ряд 4	Ряд 5
Пипетка объемом 25 мл	Пипетка объемом 10 мл	Пипетка объемом 20 мл	Пипетка объемом 50 мл	Пипетка объемом 25 мл
Стандартный р-р KOH 0,1	Стандартный р-р KOH 0,08	Стандартный р-р KOH 0,12	Стандартный р-р KOH 0,2	Стандартный р-р KOH 0,3
16,3	7,58	11,0	16,0	6,0
15,2	7,65	10,1	13,9	5,2
15,7	7,47	9,9	14,5	4,9
14,9	7,30	9,8	13,8	3,9
14,9	7,21	9,9	16,2	5,5
15,0	7,54	10,2	16,0	5,1
15,2	7,29	10,3	15,3	6,0
16,2	6,98	10,4	14,3	4,8
15,3	7,67	10,2	14,2	5,2
15,5	6,60	8,9	13,9	5,3
14,9	6,38	9,8	15,4	3,9
16,2	7,27	11,2	16,0	5,2
15,8	7,02	11,3	16,3	5,4
15,4	7,0	12,3	14,8	3,9
16,4	5,89	8,9	13,8	5,5

Ряд 6	Ряд 7	Ряд 8	Ряд 9	Ряд 10
Пипетка объемом 2 мл	Пипетка объемом 20 мл	Пипетка объемом 5 мл	Пипетка объемом 100 мл	Пипетка объемом 25 мл
Стандартный р-р KOH 0,015	Стандартный р-р KOH 0,08	Стандартный р-р KOH 0,025	Стандартный р-р KOH 0,65	Стандартный р-р KOH 0,11
9,0	13,0	13,0	9,8	12,7
7,9	15,4	12,8	9,5	13,8
7,8	15,2	12,4	9,4	13,7
7,7	15,0	12,3	7,9	12,6
8,2	16,3	12,1	9,2	13,4
8,3	13,6	11,0	8,9	13,9
7,8	15,0	11,5	9,4	13,4
8,0	15,2	11,8	8,0	13,3
9,1	15,4	11,7	9,2	13,6
8,05	13,8	11,9	8,9	12,8
8,2	14,7	12,0	9,7	13,7
7,8	14,8	12,1	8,6	13,9
8,2	13,9	13,0	9,4	13,2
8,1	13,8	12,3	9,9	13,9
7,9	13,7	12,5	9,8	13,8

Практическая работа №2

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Обработку серии измерений следует производить в следующем порядке:

1 – определить среднее арифметическое;

2 – найти среднюю квадратичную ошибку отдельного измерения;

3 – определить наибольшую возможную ошибку Δ отдельного измерения и убедиться, что среди результатов измерений нет таких, которые бы отличались от среднего арифметического более чем на Δ. Если бы такие оказались, их следует отбросить и начать обработку сначала;

4 – определить среднюю квадратичную ошибку σ_0 среднего арифметического и остальные характеристики (r_0, Δ_0, h, H).

Примеры заданий с решениями.

Обработать 16 измерений, приведенных таблице 2 и представляющих собой результаты анализа раствора на содержание в нем растворенной кислоты НА.

Находим среднее арифметическое

$$\bar{x} = \frac{1632}{16} = 102{,}0$$

В 3-й столбец заносим отклонения отдельных измерений от среднего арифметического, а в 4-ый – их квадраты. При этом $\sum(\bar{x} - x_i)$ должна быть равна нулю, что является контролем правильности вычислений.

Таблица 2 – Пример статистической обработки экспериментальных данных

Исходные данные		Первая обработка		Вторая обработка		
№ п/п	x_i, г/л	$\bar{x}$ - x_i	$(\bar{x}$ - $x_i)^2$	x_i, г/л	$\bar{x}$ - x_i	$(\bar{x}$ - $x_i)^2$
1	102	0	0	102	2,53	6,4
2	98	4	16	98	1,47	2,2
3	99	3	9	99	0,47	0,2
4	100	2	4	100	-0,53	0,3
5	97	5	25	97	2,47	6,1
6	140	-38	1444	—	—	—
7	95	7	49	95	4,47	20,0
8	100	2	4	100	-0,53	0,3
9	98	4	16	98	1,47	2,2
10	96	6	36	96	3,47	12,0
11	102	0	0	102	-2,53	6,4
12	101	1	1	101	-1,53	2,3
13	101	1	1	101	-1,53	2,3
14	102	0	0	102	-2,53	6,4
15	99	3	9	99	0,47	0,2
16	102	0	0	102	-2,53	6,4
	1632	0	1614	1492	0,15	73,7

Далее вычисляем по формуле наибольшую возможную ошибку отдельных наблюдений:

$$\Delta = 3\sigma = 3\sqrt{\frac{\sum(\bar{x}-x_i)^2}{n-1}} = 3\sqrt{\frac{1614}{16-1}} = 31$$

Сопоставляя эту ошибку с цифрами 3-го столбца таблицы, видим, что 6-ое наблюдение приходится отклонить, как недоброкачественное (38>31).

Производим в столбцах 5 – 7 табл. вычисления заново

$\bar{x}$ = 1492/15 = 99,47

$$\sigma = \sqrt{\frac{73{,}7}{14}} = 2{,}21; \qquad \Delta = 3\cdot\sqrt{\frac{3{,}73}{14}} = 6{,}6$$

Все числа 6-го столбца — меньше 6,6 и, следовательно, все 15 оставшихся наблюдений должны быть учтены.

Вычисляем среднюю квадратичную ошибку среднего арифметического по формуле

$$\sigma_0 = \sqrt{\frac{\sum_{i=1}^{n}(\bar{x}-x_i)^2}{n(n-1)}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_0 = \sqrt{\frac{73{,}7}{15\cdot 14}} \approx 0{,}6$$

Результат обработки можно записать одним из следующих способов $\bar{x}$ = 99,5 ± 0,6; или $\bar{x}$ = 99,5(1 ± 0,006); или $\bar{x}$ = 99,5 с точностью 0,6%.

Для вероятной ошибки отдельных измерений находим следующее значение

$\rho = r = 0{,}675\sigma = 0{,}675\cdot 2{,}21 = 1{,}5$

Вероятную ошибку среднего арифметического найдем по формуле

$r_0 = 0{,}675\sigma_0 \approx (2/3)\sigma_0$ $\qquad r_0 = 0{,}675\cdot 0{,}6 = 0{,}4$

Меру точности отдельных наблюдений h найдем

$$h = \sqrt{\frac{n-1}{2\sum(\bar{x}-x_i)^2}} = \sqrt{\frac{14}{2\cdot 73{,}7}} = 0{,}309$$

А мера точности среднего арифметического H будет

$$H = h\sqrt{n} = 0{,}309\sqrt{15} = 1{,}20$$

Кривая распределения вероятностей имеет уравнение

$$\varphi(x) = \frac{0{,}309}{\sqrt{\pi}} e^{-(x-99{,}5)^2\cdot(0{,}309)^2}$$

Многовариантные задачи

Для полученного ряда значений выполнить статистическую обработку данных с построением кривой распределения Гаусса.

Таблица 3 - Задание для вычисления

Ряд 1	Ряд 2	Ряд 3	Ряд 4	Ряд 5
Пипетка объемом 15 мл	Пипетка объемом 5 мл	Пипетка объемом 10 мл	Пипетка объемом 25 мл	Пипетка объемом 50мл
Стандартный р-р КОН 0,2	Стандартный р-р КОН 0,02	Стандартный р-р КОН 0,02	Стандартный р-р КОН 0,1	Стандартный р-р КОН 0,025
16,3	7,58	11,0	16,0	6,0
15,2	7,65	10,1	13,9	5,2
15,7	7,47	9,9	14,5	4,9
14,9	7,30	9,8	13,8	3,9
14,9	7,21	9,9	16,2	5,5
15,0	7,54	10,2	16,0	5,1
15,2	7,29	10,3	15,3	6,0
16,2	6,98	10,4	14,3	4,8
15,3	7,67	10,2	14,2	5,2
15,5	6,60	8,9	13,9	5,3
14,9	6,38	9,8	15,4	3,9
16,2	7,27	11,2	16,0	5,2
15,8	7,02	11,3	16,3	5,4
15,4	7,0	12,3	14,8	3,9
16,4	5,89	8,9	13,8	5,5

Ряд 6	Ряд 7	Ряд 8	Ряд 9	Ряд 10
Пипетка объемом 25 мл	Пипетка объемом 2мл	Пипетка объемом 15 мл	Пипетка объемом 50 мл	Пипетка объемом 15 мл
Стандартный р-р КОН 0,01	Стандартный р-р КОН 0,001	Стандартный р-р КОН 0,25	Стандартный р-р КОН 0,05	Стандартный р-р КОН 0,05
9,0	13,0	13,0	9,8	12,7
7,9	15,4	12,8	9,5	13,8
7,8	15,2	12,4	9,4	13,7
7,7	15,0	12,3	7,9	12,6
8,2	16,3	12,1	9,2	13,4
8,3	13,6	11,0	8,9	13,9
7,8	15,0	11,5	9,4	13,4
8,0	15,2	11,8	8,0	13,3
9,1	15,4	11,7	9,2	13,6
8,05	13,8	11,9	8,9	12,8
8,2	14,7	12,0	9,7	13,7
7,8	14,8	12,1	8,6	13,9
8,2	13,9	13,0	9,4	13,2
8,1	13,8	12,3	9,9	13,9
7,9	13,7	12,5	9,8	13,8

Практическая работа №3.

РАСПРЕДЕЛЕНИЕ ПРОМАХОВ И ИСКЛЮЧЕНИЕ ИХ ИЗ ВЫБОРКИ

Промах – погрешность, резко искажающая результат анализа, вызванная небрежностью или некомпетентностью работника, обычно легко обнаруживаемая.

Грубые ошибки или промахи возникают в результате серьезных отклонений от стандартных условий эксперимента. Их следует всячески избегать, но, к несчастью, время от времени они все-таки случаются. Причинами могут быть сбои в работе прибора, значительные загрязнения реактивов, случайные потери образца и т. д. При обнаружении источников грубых погрешностей обычно не остается ничего иного, кроме как прекратить эксперимент и начать его заново. Представляет собой отклонение, которое резко отличается по значению от других отклонений выборки и причиной которого является невнимательность или некомпетентность исполнителя.

При небрежной работе полученные результаты должны быть исключены, так как в противном случае среднее арифметическое значение и стандартная погрешность будут искажаться. В случае выборки результатов и трехсигмовой доверительной вероятности промахом считается результат, для которого отклонение $X_i - \overline{X} \geq 3$. Иногда для обнаружения в выборке промаха сначала рассчитывают значение S для всех вариант, выбрасывают из выборки результат, который при выбранной доверительной вероятности оказался промахом, снова рассчитывают S, исключают варианты, которые не попадают в новую выборку, и т. д., пока промахи не будут отсутствовать.

Рассмотренный путь выявления промахов длительный. Проще выявить промахи можно, используя Q-тест (критерий). Варианты выборки располагают для этого в порядке возрастания и путем деления разности подвергаемой сомнению и соседней с ней вариант на диапазон выборки находят расчетное значение Q_P

$$Q_p = \frac{X_{i+1} - X_i}{\omega}$$

которое затем сравнивают с табличным значением $Q_т$ (таблица 4). Если $Q_P \geq Q_т$, то проверяемый результат является промахом и его отбрасывают. Для выборки из трех вариант проверку начинают с наименьшего значения. Если при этом наименьшее значение подлежит исключению, то дополнительно выполняют еще одно -два определения, применяют Q-тест к новой выборке при отсутствии промаха и рассчитывают среднее арифметическое. При $n > 3$ первой проверяют наибольшую варианту. Если она является промахом, то ее исключают, затем находят диапазон для новой выборки и для нее проверяют уже наименьшее значение и т. д.

Примеры заданий с решениями.

Проверить наличие промахов в выборке результатов при определении содержания Na_2CO_3 в растворе соды (100 мл) путем прямого титрования его аликвотных частей раствором HCl. Результаты (в г): 0,2031; 0,2033; 0,2015; 0,2048; 0,2020.

Таблица 4 Табличные значения $Q_т$.

Число вариант	Табличные значения $Q_т$ при вероятности		
	0,90	0,95	0,99
3	0,94	0,98	0,99
4	0,76	0,85	0,93
5	0,64	0,73	0,82
6	0,56	0,64	0,74
7	0,51	0,59	0,68
8	0,47	0,54	0,63
9	0,44	0,51	0,60
10	0,41	0,48	0,57

Решение. Располагаем результаты в порядке их возрастания: 0,2015; 0,2020; 0,2031; 0,2033; 0,2048, находим диапазон выборки; ω = 0,2048 — 0,2015 = 0,0033 и проверяем максимальную варианту, так как число вариант в выборке больше трех:

Q_P = (0,2048 — 0,2033) /0,0033 = 0,0015/0,0033 = 0,45.

При P = 0,95 $Q_т$. = 0,73 в случае выборки из пяти вариант. Так как $Q_P < Q_т$, то максимальное значение сохраняется и, следовательно, промахи в выборке отсутствуют.

Многовариантная задача

Для полученного ряда проверить наличие промахов в выборке результатов (таблица 5)

Таблица 5– Значения, подлежащие проверке на наличие промахов

№	Ряд 1	Ряд 2	Ряд 3	Ряд 4	Ряд 5	Ряд 6	Ряд 7	Ряд 8	Ряд 9	Ряд 10
1	5,72	5,72	5,09	5,19	5,72	5,14	5,45	5,33	5,72	5,21
2	6,85	6,85	6,93	6,54	6,95	6,83	6,88	5,87	6,25	6,21
3	15,9	15,3	15,1	15,5	15,5	15,6	15,6	15,3	15,3	15,0
4	5,53	5,53	5,72	5,10	5,73	5,68	5,59	5,67	5,61	5,14
5	51,5	52,5	51,4	52,3	53,1	52,5	51,7	55,0	51,0	54,1
6	25,5	25,7	25,5	26,8	27,3	25,7	25,6	25,7	20,1	22,3
7	75,8	75,8	76,8	75,2	75,9	75,6	75,8	75,3	73,5	72,8
8	0,92	0,91	0,88	0,96	0,92	0,93	0,94	0,85	0,87	0,90
9	2,54	2,56	2,87	2,98	2,64	2,31	2,57	2,68	2,98	2,16
10	4,47	4,57	4,63	4,25	4,15	4,68	4,98	4,06	4,35	4,67
11	6,53	6,53	6,72	6,10	6,63	6,68	6,69	6,67	5,91	6,94
12	0,62	0,61	0,68	0,66	0,66	0,68	0,74	0,65	0,67	0,60

Практическая работа №4.

ОБРАБОТКА РЕЗУЛЬТАТОВ НЕРАВНОТОЧНЫХ НАБЛЮДЕНИЙ ПРИ РАЗНОМ ЧИСЛЕ ИЗМЕРЕНИЙ В РЯДАХ, НО ОДИНАКОВОЙ ТОЧНОСТИ КАЖДОГО ОТДЕЛЬНОГО НАБЛЮДЕНИЯ И НАБЛЮДЕНИЙ С РАЗНОЙ ТОЧНОСТЬЮ ОТДЕЛЬНЫХ ИЗМЕРЕНИЙ

Если каждый ряд содержит одинаковое число измерений, то результаты обработки рядов будут равноточны. Если же число измерений в рядах не одинаковое, то результаты обработки рядов будут не равноточны. Это следует из того, что в формулу средней квадратичной ошибки арифметического среднего

$$\sigma_0 = \sigma\sqrt{n}$$

входит число измерений n.

Сущность обработки рядов неравноточных измерений заключается в том, что после введения некоторых коэффициентов, называемых «весами», обработку неравноточных измерений производят так же как и равноточных.

Примеры заданий с решениями.

Концентрация раствора кислоты (г/кг) определялась четырьмя группами измерений, представленными в таблице 6.

Таблица 6 - Концентрация кислоты, определенная в 4 группах

N п/п	[HA], г/кг			
	1-я группа	2-я группа	3-я группа	4-я группа
1	107,10	107,57	107,51	107,42
2	107,68	107,45	107,51	107,00
3	107,45	107,07	107,16	
4	107,62	107,35	107,48	
5	107,68	107,17		
6	107,08	107,46		
7	107,44			
8	107,28			
Σ	859,33	644,07	429,72	214,42
n	8	6	4	2
$\bar{x}$	107,416	107,345	107,430	107,210

Для каждой группы измерений определены средние арифметические значения. В рассматриваемом примере мы имеем 4-е средних арифметических с весами: 107,416 с весом 8; 107,345 с весом 6; 107,430 с весом 4; 107,210 с весом 2

Окончательная средняя арифметическая, называемая общей арифметической серединой или взвешенной средней, будет

$$\bar{x} = \frac{107{,}416 \cdot 8 + 107{,}345 \cdot 6 + 107{,}430 \cdot 4 + 107{,}210 \cdot 2}{8 + 6 + 4 + 2} = 107{,}38$$

Дальнейшую обработку проводить как в практической работе №2.

Многовариантная задача

Для полученных значений, выполненных с одинаковой точностью, но отличающихся числом измерений в каждом ряду, провести статистическую обработку (таблица 7, таблица 8).

Таблица 7 – Условия для титрования

№ варианта	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
V пипетки, мл	0,5	50	25	15	10	20	100	0,1	2	5
С раствора, моль/л	0,01	0,02	0,2	0,05	0,25	0,54	0,15	0,005	0,033	2,0

Таблица 8 – Значения объемов, полученных при титровании

№	Ряд 1	Ряд 2	Ряд 3	Ряд 4	Ряд 5	Ряд 6	Ряд 7	Ряд 8	Ряд 9	Ряд 10
1	5,7	5,7	5,4	5,1	5,7	5,1	5,4	5,3	5,7	5,2
2	5,8	5,8	5,9	5,5	5,9	5,8	5,8	5,8	5,2	5,2
3	5,3	5,3	5,1	5,5	5,5	5,6	5,6	5,3	5,3	5,3
4	5,5	5,5	5,7	5,1	5,7	5,6	5,5	5,6	5,1	5,1
5	5,5	5,5	5,5	5,3	5,3	5,5	5,5	5,3	5,5	5,1
6	5,5	5,7	5,5	5,8	5,3	5,7	5,6	5,7	5,1	
7	5,8	5,8	5,9	5,5	5,9	5,6	5,8	5,3	5,1	
8	5,3	5,3	5,1	5,0	5,6	5,3	5,3	5,3	5,5	
9	5,7	5,5	5,3	5,3	5,9	5,6	5,4	5,5		
10	5,5	5,7	5,5	5,9	5,9	5,7	5,5	5,5		
11	5,7	5,4	5,4	5,5	5,6		5,4	5,9		
12	5,8	5,8	5,9	5,0	5,1		5,8	5,1		
13	5,9	5,9	5,9	5,1	5,7		5,9			
14	5,5	5,5	5,7	5,7			5,6			
15	5,3	5,0	5,5	5,7			5,6			
16	5,5	5,4	5,7	5,5			5,0			
17	5,8	5,8	5,9				5,8			
18	5,3	5,3	5,1				5,6			
19	5,9	5,9					5,9			
20	5,2	5,2					5,2			
21	5,8	5,8					5,0			
22	5,8						5,8			
23	5,9						5,0			
24	5,2									
25	5,3									

Практическая работа №5

ОБРАБОТКА РЕЗУЛЬТАТОВ НЕРАВНОТОЧНЫХ НАБЛЮДЕНИЙ ПРИ РАЗНОМ ЧИСЛЕ ИЗМЕРЕНИЙ В РЯДАХ, НО ОДИНАКОВОЙ ТОЧНОСТИ КАЖДОГО ОТДЕЛЬНОГО НАБЛЮДЕНИЯ

Если даны два ряда измерений, но измерения каждого ряда произведены приборами разной точности, то арифметические средние будут не равноточны даже в том случае, если бы число измерений в каждом ряду было бы одинаковым, так как в формулу

$$\sigma_0 = \sqrt{\frac{\sum(\bar{x} - x_i)^2}{n(n-1)}} = \frac{\sigma}{\sqrt{n}}$$

входит число σ.

Из формулы, выражающей среднюю квадратичную ошибку σ_0 среднего арифметического через среднюю квадратичную ошибку σ отдельного измерения и число n измерений, можно получить

$$n = \sigma^2 / \sigma_0{}^2$$

Если отдельные измерения обладают одинаковой точностью, то средним арифметическим отдельных групп измерений следует приписать «веса», обратно пропорциональные квадратам их средних квадратичных ошибок. Так как средняя квадратичная, вероятная и наибольшая возможная ошибки пропорциональны друг другу, то в качестве «весов» средних арифметических можно взять числа обратно пропорциональные квадратам любых этих ошибок.

Пример заданий с решениями.

1. Угол смачивания при осуществлении процесса аэрации флотируемой пульпы измерен в трех группах наблюдений с указанием средней квадратичной ошибки:

$30^0\ 15'\ 30'' \pm\ 40''$ $30^0\ 15'\ 15'' \pm\ 20''$ $30^0\ 15'\ 20'' \pm\ 10''$

Найти общую арифметическую середину.

Решение. В качестве «весов» каждой группы наблюдений принимаем числа, обратно пропорциональные числам 40^2, 20^2 и 10^2, т.е.

$$g_1 : g_2 : g_3 = 1/1600 : 1/400 : 1/100$$

Следовательно, $g_1 = 1$; $g_2 = 4$; $g_3 = 16$ и взвешенное среднеарифметическое равно

$$\bar{x} = 30^0 15' + \frac{30''\cdot 1 + 15''\cdot 4 + 20''\cdot 16}{1+4+16} = 30^0 15' 19{,}5''$$

2. Выполнено шесть серий измерений значения размера и получены следующие результаты: значение размера в серии 20,617;

20,666; 20,643; 20,635; 20,629 и 20,654; среднее квадратичное отклонение размера в i-й серии равны соответственно 32; 24; 18; 20; 16; 16. Найти среднюю взвешенную.

Решение. В качестве «весов» каждой группы наблюдений принимаем числа, обратно пропорциональные квадратам средне квадратичных отклонений, т.е.

$$g_1: g_2: g_3: g_4:g_5: g_6= (1/1024): (1/576): (1/324):(1/400):(1/256):(1/256)$$

$$\bar{x} = \frac{20{,}617 \cdot 8100 + 20.666 \cdot 14400 + 20.643 \cdot 25600 + 20.635 \cdot 20736 + 20.629 \cdot 32400 + 20.645 \cdot 32400}{8100 + 14400 + 25600 + 20736 + 32400 + 32400} =$$

$=20{,}639$.

Задачи

1. Определить среднее значение следующих групп измерений величины краевого угла

20^0 11’ 30’’ ± 40’’
20^0 10’ 78’’ ± 20’’
20^0 11’ 20’’ ± 10’’
20^0 11’ 10’’ ± 15’’
20^0 11’ 40’’ ± 25’’
20^0 11’ 20’’ ± 15’’

Многовариантная задача

При измерении времени, за которое тело проходит некоторый постоянный путь, проведено 6 серий измерений. После обработки отсчетов каждой серии были получены следующие результаты (таблица 9). Определить среднюю арифметическую взвешенную каждого ряда

Таблица 9–Значения для вычисления

№ варианта	№ измерения					
	1	2	3	4	5	6
1	0.2±0.1	0.3±0.2	0.4±0.2	0.18±0.15	0.22±0.21	0.32±0.15
2	0.3±0.1	0.7±0.2	0.5±0.2	0.28±0.15	0.42±0.21	0.62±0.15
3	0.9±0.1	0.8±0.2	0.7±0.2	0.48±0.17	0.52±0.21	0.62±0.15
4	0.6±0.1	0.5±0.2	0.7±0.2	0.88±0.15	0.92±0.21	0.72±0.15
5	0.7±0.1	0.5±0.6	0.6±0.25	0.68±0.19	0.62±0.21	0.62±0.15
6	0.8±0.3	0.7±0.2	0.9±0.2	0.88±0.15	0.82±0.29	0.82±0.15
7	0.5±0.1	0.3±0.5	0.7±0.25	0.58±0.18	0.52±0.21	0.72±0.15
8	0.2±0.1	0.7±0.2	0.3±0.26	0.38±0.15	0.22±0.26	0.32±0.12
9	0.7±0.5	0.3±0.4	0.6±0.27	0.28±0.14	0.28±0.21	0.37±0.15
10	0.25±0.1	0.38±0.2	0.47±0.23	0.28±0.15	0.22±0.27	0.32±0.11
11	0.72±0.1	0.93±0.6	0.44±0.26	0.28±0.12	0.22±0.21	0.52±0.15
12	0.82±0.7	0.53±0.2	0.43±0.27	0.68±0.15	0.72±0.24	0.36±0.15
13	0.5±0.1	0.32±0.2	0.44±0.26	0.58±0.13	0.25±0.21	0.62±0.15
14	0.26±0.1	0.34±0.3	0.74±0.27	0.68±0.15	0.42±0.21	0.62±0.16
	5±1	6±1	4±3	5±2	7±1	9±4

Практическая работа №6

КРИТЕРИЙ ФИШЕРА И ЕГО ИСПОЛЬЗОВАНИЕ В ХИМИЧЕСКОЙ ПРАКТИКЕ

Во многих случаях важно сравнить изменчивость или «размах» двух или большего числа выборок данных. Проверку равноточности выборок можно провести их попарным сравнением при помощи критерия Фишера (F-критерий)

Рассматриваемый ниже критерий применяется при сравнении точности двух рядов измерений, при проверке устойчивости технологического процесса и в других подобных вариантах.

Функция F есть отношение выборочных дисперсий

$$F = \sigma_1^2 / \sigma_2^2$$

Значение этой функции для уровней значимости 0,05 и 0,01 табулированы при соответствующих степенях свободы каждой из двух дисперсий (таблица 10). При сравнении двух дисперсий обычно в числителе критерия F содержится большая дисперсия.

Таблица 10 – Теоретические значения критерия Фишера (F_T)

f_2	Значения F_T при f_1				
	2	3	4	5	6
P=0,95					
2	19,00	19,16	19,25	19,30	19,33
3	9,55	9,28	9,12	9,01	8,94
4	6,94	6,59	6,39	6,26	6,16
5	5,79	5,41	5,19	5,05	4,95
6	5,14	4,76	4,53	4,39	4,28
P=0,99					
2	99,00	99,17	99,25	99,30	99,33
3	30,81	29,46	28,71	28,24	27,91
4	18,00	16,69	15,98	15,52	15,21
5	13,27	12,06	11,39	10,97	10,67
6	10,92	9,78	9,15	8,75	8,47

По таблице 10 находят теоретическое значение $F_т$. Если $F_p < F_{тг}$, выборки принадлежат одной генеральной совокупности и, следовательно, равноточны. Число степеней свободы большей дисперсии n_1— 1 и, соответственно, для меньшей дисперсии n_2— 1- Обращает на себя внимание, что $F_т$ сильнее зависит от числа степеней свободы выборки с меньшей стандартной погрешностью S_2, т. е. имеющей большую точность, а также его значительное различие для двух приведенных доверительных вероятностей при одинаковых степенях свободы. Доверительная вероятность P=0,95 является более жесткой, так как исключает из генеральной совокупности меньшую разницу вариант. Поэтому равноточность двух выборок, полученная при P=0,99, может оказаться недоказанной при P=0,95

Пример задачи с решением

При определении содержания хлора в полимерном материале использованы два метода анализа, результаты которых приведены ниже (в %)

Метод А: $X_А$ 27,5 27,0 27,3 27,6 27,8

Метод Б: $X_Б$ 27,9 26,5 27,2 26,3 27,0 27,4 27,3 26,8

Решение. Для каждой выборки находим дисперсию

$$\sigma_A{}^2 = \frac{\sum x_A{}^2 - \frac{\left(\sum x_A\right)^2}{n}}{n-1} = 0{,}093; \quad \sigma_B{}^2 = \frac{\sum x_B{}^2 - \frac{\left(\sum x_B\right)^2}{n}}{n-1} = 0{,}266.$$

Критерий Фишера равен F = 0,266/0,093 = 2,86. В таблице значений F находим F = 6,09 применительно к вероятности 0,05 при 7 степенях свободы для $\sigma_1{}^2$ и 4 степенях свободы для $\sigma_2{}^2$.

Так как полученное значение F = 2,86 значительно меньше табличного, то наше предположение об отсутствии различий между методами анализа в отношении воспроизводимости при данных условиях не опровергается. Различия в воспроизводимости носят случайных характер и обе дисперсии являются приближенными оценками одной и той же общей для обеих выборок дисперсий генеральной совокупности.

Если расчетное значение F больше табличного, имеются основания сомневаться в том, что эти две дисперсии соответствуют одной и той же совокупности.

Многовариантная задача.

Для двух рядов измерений, выбранных преподавателем из выборки рядов многовариантной задачи практической работы №2 «Статистическая обработка экспериментальных данных» проверить принадлежность к одной совокупности.

Практическая работа №7.

КРИТЕРИЙ СТЬЮДЕНТА И ЕГО ИСПОЛЬЗОВАНИЕ В ХИМИЧЕСКОЙ ПРАКТИКЕ

В дисперсионном анализе наравне с F-критерием применяется и другой критерий, который называется t-критерий Стьюдента. Критерий Стьюдента применяется в том случае, когда требуется дать ответ:

отличаются ли достоверно, т.е. надежно, результаты одной группы от результатов другой группы.

Оценивается степень расхождения средних арифметических показателей двух групп данных (M_1 и M_2) относительно дисперсии σ^2, т.е разброса индивидуальных данных, рассчитанной применительно к этим двум группам, где количество членов соответствует N_1 и N_2. В том случае, когда выборки примерно равны по численности используется эта формула.

$$t = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

В случае, когда выборки отличаются по своим размерам, применяется другая – более сложная формула расчета критерия:

$$t = \frac{M_1 - M_2}{\sigma_d}, \quad \sigma_d = \sqrt{\frac{\sigma_1^2 \cdot (N_1 - 1) + \sigma_2^2 \cdot (N_2 - 1)}{N_1 + N_2 - 2}} \cdot \sqrt{\frac{1}{N_1} \cdot \frac{1}{N_2}}$$

Полученное значение t-критерия Стьюдента необходимо правильно интерпретировать. Для этого нам необходимо знать количество исследуемых в каждой группе (n_1 и n_2). Находим число степеней свободы f по следующей формуле:

$$f = (n_1 + n_2) - 2$$

После этого определяем критическое значение t-критерия Стьюдента для требуемого уровня значимости (например, p=0,05) и при данном числе степеней свободы f по таблице 11.

Сравниваем критическое и рассчитанное значения критерия:

1) если рассчитанное значение t-критерия равно или больше найденного по таблице, делаем вывод о статистической значимости различий между сравниваемыми величинами.

2) если значение рассчитанного t-критерия меньше табличного, значит различия сравниваемых величин статистически не значимы.

Пример задачи с решением

Для изучения эффективности нового препарата железа были выбраны две группы пациентов с анемией (численность первой группы составила 34, а второй - 40 пациентов). В первой группе пациенты в течение двух недель получали новый препарат, а во второй группе - получали плацебо. После измерения средний уровень гемоглобина составил 115,4±1,2 г/л в первой группе, и 103,7±2,3 г/л во второй (данные представлены в формате M±m, сравниваемые совокупности имеют нормальное распределение). Сделать вывод о статистической значимости полученных различий и эффективности нового препарата железа.

Таблица 11- Значения -критерия Стьюдента при p=0.05

Число степеней свободы, f	Значение t-критерия	Число степеней свободы, f	Значение t-критерия	Число степеней свободы, f	Значение t-критерия
1	12.706	24	2.064	56-57	2.003
2	4.303	25	2.060	58-59	2.002
3	3.182	26	2.056	60-61	2.000
4	2.776	27	2.052	62-63	1.999
5	2.571	28	2.048	64-65	1.998
6	2.447	29	2.045	66-67	1.997
7	2.365	30	2.042	68-69	1.995
8	2.306	31	2.040	70-71	1.994
9	2.262	32	2.037	72-73	1.993
10	2.228	33	2.035	74-75	1.993
11	2.201	34	2.032	76-77	1.992
12	2.179	35	2.030	78-79	1.991
13	2.160	36	2.028	80-89	1.990
14	2.145	37	2.026	90-99	1.987
15	2.131	38	2.024	100-119	1.984
16	2.120	40-41	2.021	120-139	1.980
17	2.110	42-43	2.018	140-159	1.977
18	2.101	44-45	2.015	160-179	1.975
19	2.093	46-47	2.013	180-199	1.973
20	2.086	48-49	2.011	200	1.972
21	2.080	50-51	2.009	∞	1.960
22	2.074	52-53	2.007		
23	2.069	54-55	2.005		

Решение: для оценки значимости различий используем t-критерий Стьюдента, рассчитываемый как разность средних значений, поделенная на сумму квадратов ошибок:

$$t = \frac{115,4 - 103,7}{\sqrt{1,2^2 + 2,3^2}}$$

После выполнения расчетов, значение t-критерия оказалось равным 4,51. Число степеней свободы равно 72.

Сравниваем полученное значение t-критерия Стьюдента 4,51 с критическим при p=0,05 значением, указанным в таблице 11: 1,993. Так как рассчитанное значение критерия больше критического, делаем вывод о том, что наблюдаемые различия статистически значимы (уровень значимости p<0,05).

Многовариантная задача.

Для двух рядов измерений, предоставленных преподавателем проверить отличаются ли достоверно, т.е. надежно, результаты одной группы от результатов другой группы.

Практическая работа №8.

ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА С ИСПОЛЬЗОВАНИЕМ СПОСОБА НАИМЕНЬШИХ КВАДРАТОВ

Метод наименьших квадратов — математический метод, применяемый для решения различных задач, основанный на минимизации суммы квадратов отклонений некоторых функций от искомых переменных. Применяется в тех случаях, когда искомые величины x_1, x_2, ... , x_n не могут быть измерены непосредственно, но могут быть представлены в виде явных функций измеренных величин. Прежде чем обрабатывать данные с применением метода наименьших квадратов, необходимо выявить промахи и исключить их из числа рассматриваемых результатов выборочной совокупности.

Пример задачи с решением

Решить методом наименьших квадратов систему уравнений:

$5{,}7x - 4{,}4y = -1{,}4$

$3{,}6x + 1{,}9y = 9{,}1$

$4{,}1x + 8{,}2y = 22{,}7$

$8{,}4x + 3{,}1y = 18{,}3$

$11{,}0x + 5{,}2y = 22{,}0$

Для составления нормальных уравнений и для оценки точности корней составляем следующую таблицу 12.

Полученные нормальные уравнения

$253{,}9x + 98{,}5y = 557{,}6$

$98{,}5x + 126{,}8y = 401{,}6$

Имеют корни $x_0 = 1{,}381$; $y_0 = 2{,}090$.

Следовательно,

$\mu_x = \sqrt{0{,}10/177(5-2)} = 0{,}014$;

$$\mu_y = \sqrt{0{,}10/89 \cdot 3} = 0{,}019 \cdot$$

Таблица 11 -Рекомендуемая форма записи расчета способом наименьших квадратов

Исходные данные					Первое нормальное уравнение			
1	2	3	4	5	6	7	8	9
N п/п	a_i	b_i	f_i	$\sigma_i = a_i + b_i + f_i$	a_i^2	$a_i b_i$	$a_i f_i$	$a_i \sigma_i$
1	5,7	-4,4	-1,4	0,1	32,5	-25,1	-8,0	-0,6
2	3,6	1,9	9,1	14,6	13,0	6,8	32,8	52,6
3	4,1	8,2	22,7	35,0	16,8	33,6	93,1	143,5
4	8,4	3,1	18,3	29,8	70,6	26,0	153,7	250,3
5	11,0	5,2	26,0	42,2	121,0	57,2	286,0	464,2
—	32,8	14,0	74,7	$\frac{121{,}5}{121{,}5}$	253,9	98,5	557,6	$\frac{910{,}0}{910{,}0}$

Второе нормальное уравнение				Оценка точности			
10	11	12	13	14	15	16	17
$b_i a_i$	b_i^2	$b_i f_i$	$b_i \sigma_i$	$a_i x_0$	$b_i y_0$	$[\varepsilon_i]$	ε^2
-25,1	19,4	6,2	0,4	7,88	-9,21	0,07	0,00
6,8	3,6	17,3	27,7	4,97	3,97	0,16	0,03
33,6	67,2	186,1	287,0	5,66	17,15	0,11	0,01
26,0	9,6	56,7	92,4	11,6	6,48	0,22	0,05
57,2	27,0	135,2	219,6	15,2	10,88	0,88	0,01
98,5	126,8	401,5	$\frac{626{,}8}{626{,}9}$	—	—	—	0,10

Многовариантная задача.

Для функции, заданной таблично преподавателем, найти линейную, вычисленную по методу наименьших квадратов.

Практическая работа №9.

РАСЧЕТ РЕГРЕССИОННЫХ УРАВНЕНИЙ

Когда в простейшем случае две переменные величины X и Y зависят друг от друга так, что каждому значению одной из них соответствует вполне определенное одно или несколько значений другой, то между ними имеется функциональная связь, выраженная уравнениями

$$Y = f(X) \qquad \text{или} \qquad X = \varphi(Y),$$

то в этом случае связь, существующая между переменными X и Y, называется *функциональной связью.*

Если приходится иметь дело с такими переменными величинами, между которыми существует зависимость, но эта зависимость не является вполне определенной: каждому значению одной из величин (например, X) соответствует некоторая совокупность значений другой (например, Y), причем распределение Y меняется определенным образом при изменении X. В этом случае связь, существующая между переменными X и Y, называется *корреляционной связью*.

Во всех таких случаях мы имеем парные значения соответствующих друг другу величин X и Y: (X_1,Y_1), (X_2,Y_2), (X_3,Y_3), ..., (X_n,Y_n). Если изобразить их на диаграмме в координатах X – Y, то получится система точек, где каждому значению X не соответствует вполне определенное значение Y, но очевидна тенденция к расположению точек определенным образом, полосой, что дает возможность установить корреляцию X и Y

$$\overline{Y} = aX + b,$$

являющаяся уравнением прямой AB и приближенно отражающая связь между X и Y. Линия AB называется линией *регрессии* Y по X.

Пример задачи с решением

Сырье, поступающее на завод из ближайшего карьера, содержит два полезных компонента – минералы А и Б. При этом в партиях сырья с повышенным содержанием А обычно обнаруживается и более высокое содержание Б, так что имеются основания ожидать, что эти величины находятся в связи друг с другом (таблица 13). Найти коэффициент корреляции.

Таблица 13 - Анализы 10 образцов сырья, поступившего в разное время из разных мест карьера, приведены в таблице

№ образца	1	2	3	4	5	6	7	8	9	10
% А	67	54	72	64	39	22	58	43	46	34
% Б	24	15	23	19	16	11	20	16	17	13

Решение. Обозначим процент минерала А через X, процент минерала Б – через Y. Имея в виду для вычисления r воспользоваться формулой , заполняем следующую таблицу 14.

Последовательно находим:

$$\overline{X} = \frac{499}{10} = 49{,}9; \qquad \overline{Y} = \frac{174}{10} = 17{,}4$$

$$\delta_x = \sqrt{\frac{27175}{10} - (49{,}9)^2} = 15{,}09; \quad \delta_y = \sqrt{\frac{3182}{10} - (17{,}4)^2} = 3{,}93 \qquad r = \frac{9288/10 - 49{,}9 \cdot 17{,}4}{15{,}09 \cdot 3{,}93} = 0{,}92$$

Таблица 14 – Расчетные значения по 10 образцам сырья

X	Y	X^2	Y^2	XY	X + Y	$(X+Y)^2$
67	24	4489	576	1608	91	8281
54	15	2916	225	810	69	4761
72	23	5184	529	1656	95	9025
64	19	4096	361	1216	83	6889
39	16	1521	256	624	55	3025
22	11	484	121	242	33	1089
58	20	3364	400	1160	78	6084
43	16	1849	256	688	59	3481
46	17	2116	289	782	63	3969
34	13	1156	169	442	47	2209
499	174	27175	3182	9228		48813

Полученный коэффициент корреляции довольно высок (r = 0,92), что указывает на наличие тесной связи между X и Y, т.е. между содержанием минералов А и Б в сырье.

Найдем уравнения регрессии, которые позволяют вычислять наиболее вероятное содержание одного из минералов, если известно содержание другого. Определим регрессию Y по X. Коэффициент регрессии Y по X:

$$r\left(\frac{\delta_y}{\delta_x}\right) = 0{,}92 \cdot \frac{3{,}93}{15{,}09} = 0{,}24$$

Уравнение регрессии Y по X:

$$Y = 17{,}4 + 0{,}24 \cdot (X - 49{,}9) \quad \text{или} \quad Y = 5{,}42 + 0{,}24 \cdot X$$

Сопоставим вычисленные по этому уравнению $Y_{выч}$ с заданными вначале (таблица 15).

Таблица 14 – Расчетные значения

X	67	54	72	64	39	22	58	46	34
Y	24	15	23	19	16	11	20	17	13
$Y_{выч}$	22	18	23	21	15	11	19	16	14

Хорошо видно, что совпадение получилось удовлетворительное.

Многовариантная задача.

Определить коэффициент корреляции для значений, заданных преподавателем.

Библиографический список

1. Дядик В.Ф.Статистические методы контроля и управления: учебное пособие / В.Ф. Дядик, С.А. Байдали, Т.А. Байдали; Томский

политехнический университет. – Томск: Изд-во Томского политехнического университета, 2011. – 144 с.

2. Майстренко, А.В. Численные методы расчёта, моделирования и проектирования технологических процессов и оборудования: учебное пособие / А.В. Майстренко, Н.В. Майстренко. – Тамбов: Изд-во ФГБОУ ВПО «ТГТУ», 2011. – 144 с.

3. Кравченко Н.С. Методы обработки результатов измерений и оценки погрешностей в учебном лабораторном практикуме: учебное пособие / Н.С. Кравченко, О.Г. Ревинская; Национальный исследовательский Томский политехнический университет. - Томск: Изд-во Томского политехнического университета, 2011. - 88 с.

4. Валов, А.В. Численные методы решения уравнений для инженеров: Учебное пособие / А.В. Валов. - Челябинск: Издательский центр ЮУрГУ, 2012. - 110 с.

5. Рейзлин В.И. Численные методы оптимизации: учебное пособие / В.И. Рейзлин; Томский политехнический университет. - Томск: Изд-во ТПУ, 2011. - 105 с.

6. О. Б. Гладких, О. Н. Прокуратова Введение в численные методы: Учебно-методическое пособие. – Елец: Изд. ЕГУ им. И.А. Бунина, 2008. – 140 с.

7. Сулейманов Р.Р. Численные методы в системе MATHCAD: Лабораторный практикум. - Уфа: Башкирский институт развития образования, 2007. - 42 с.