

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

ПРАКТИКУМ ПО ИНФОРМАТИКЕ

**Статистическая обработка
химического эксперимента
средствами электронных таблиц**

Учебно-методическое пособие

Составители:
И.В. Протасова, И.В. Нечаев

Воронеж
Издательский дом ВГУ
2017

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Выборочная и интервальная статистика

Когда возможно повторение экспериментальной оценки какой-либо величины, оценки истинного значения измеряемой величины могут быть получены с помощью аппарата математической статистики. Обработка результатов эксперимента в химии, заключается в применении методов математической статистики для оценки зависимости значений различных физико-химических величин (свойства соединений, параметры химических и химико-технологических процессов и др.), характеризующих изучаемые объекты, процессы от одного или нескольких изменяемых параметров (например, температуры, давления и др.). Обработка результатов эксперимента включает определение точности данных, полученных при его проведении, оценки воспроизводимости опыта.

Результаты измерений обычно содержат случайные ошибки, поэтому статистические оценки выполняют только при наличии серии измерений – случайной выборки. Для оценки измеряемого значения какой-либо величины от изменяемых параметров по данным выборки рассчитывают выборочные параметры. Значения случайной выборки, как правило, подчиняются нормальному закону распределения случайных величин, конкретный вид которого определяют два параметра – выборочное среднее и выборочная дисперсия.

Точность получаемых оценок устанавливают с помощью статистических критериев Стьюдента (t-критерий), Фишера (F-критерий) и т. д. При этом количественными мерами служат вероятность β и уровень значимости статистического критерия $p = 1 - \beta$.

Неотъемлемой частью любого измерения является погрешность. **Погрешность** – количественная характеристика неопределенности, или неод-

$Gp = \frac{\max(s_j^2)}{\sum_{j=1}^N s_j^2}$ есть расчетное значение критерия Кохрена. Критическое, предельно допустимое, значение критерия Кохрена **Gt**, определяется из справочных таблиц по общему числу оценок дисперсий (т.е. числу серий опытов N) и числу степеней свободы $f = k - 1$, где k - число параллельных измерений в опыте [4-5]. Если выполняется условие **Gp** ≤ **Gt**, то опыты считаются воспроизводимыми.

Таблица 1.

Некоторые статистические функции в Excel и Calc

функция Excel (рус.)	Функция Calc (англ.)	Назначение
ДИСП	VAR	Возвращает дисперсию по выборке
ДИСПР	VARP	Возвращает дисперсию для генеральной совокупности
ДОВЕРИТ	CONFIDENCE	Возвращает доверительный интервал для среднего генеральной совокупности
КВАЛПРОТКЛ	DEVSQ	Возвращает сумму квадратов отклонений точек данных от их среднего
КВПИРСОН	RSQ	Возвращает квадрат коэффициента корреляции Пирсона
КОВАР	COVAR	Возвращает ковариацию, т. е. среднее произведений отклонений для каждой пары точек данных
КОРРЕЛ	CORREL	Возвращает коэффициент корреляции

ЛГРФПРИБЛ	LOGEST	Возвращает экспоненциальное уравнение регрессии
ЛИНЕЙН	LINEST	Возвращает линейное уравнение регрессии
МАКС	MAX	Возвращает максимальное значение из списка аргументов
МЕДИАНА	MEDIAN	Возвращает медиану заданного набора чисел
МИН	MIN	Возвращает наименьшее значение в списке аргументов
МОДА	MODE	Возвращает наиболее часто встречающееся значение набора данных
НАИБОЛЬШИЙ	LARGE	Возвращает k-е наибольшее значение из множества данных
НАИМЕНЬШИЙ	SMALL	Возвращает k-е наименьшее значение во множестве данных
НАКЛОН	SLOPE	Возвращает наклон линии линейной регрессии
ОТРЕЗОК	INTERCEPT	Возвращает отрезок, отсекаемый на оси линией линейной регрессии
ПИРСОН	PEARSON	Возвращает коэффициент корреляции Пирсона
ПРЕДСКАЗ	FORECAST	Возвращает предсказанное значение функции в данной точке на основе уравнение линейной регрессии
РАНГ	RANK	Возвращает ранг числа в списке чисел Ранг числа — его величина относительно других значений в списке

РОСТ	GROWTH	Возвращает значения в соответствии с экспоненциальным уравнением регрессии
СРЗНАЧ	AVERAGE	Возвращает среднее (арифметическое) значение
СРОТКЛ	AVEDEV	Возвращает среднее абсолютных значений отклонений точек данных от среднего
СТАНДОТКЛОН	STDEV	Возвращает стандартное отклонение по выборке
СТАНДОТКЛОНП	STDEVP	Возвращает стандартное отклонение по генеральной совокупности
СТЮДРАСПОБР	TINV	Возвращает обратное распределение Стьюдента для заданного числа степеней свободы
СЧЁТ	COUNT	Возвращает количество чисел в списке аргументов
СЧЁТЗ	COUNTA	Возвращает количество непустых значений в списке аргументов
ТЕНДЕНЦИЯ	TREND	Возвращает значения в соответствии с линейным уравнением регрессии
ЧАСТОТА	FREQUENCY	Возвращает распределение частот в виде вертикального массива
ЭКСЦЕСС	KURT	Возвращает эксцесс множества данных

Часто в практических задачах приходится определять интервалы отклонения оценочных значений случайной величины от ее истинного значения с разной вероятностью. Т.е. для случайной величины, за истинное зна-

чение которой принято ее среднее значение, найденное из опыта, требуется найти максимальную погрешность ε_β с заданной вероятностью β :

$$\bar{x} - \varepsilon_\beta < \bar{x} < \bar{x} + \varepsilon_\beta.$$

Полученный интервал называют доверительным, вероятность β – доверительной вероятностью, а его границы $(\bar{x} \pm \varepsilon_\beta)$ – доверительными границами. На практике величину доверительной вероятности берут 0,90; 0,95; 0,99. Расчет доверительного интервала зависит от объема выборки – k . Если объем выборки больше 30 ($k > 30$), то предполагают, что случайная величина подчиняется нормальному распределению, и расчет доверительного интервала можно проводить с использованием стандарта среднего

$$S_{cp} = \frac{\sqrt{S^2}}{\sqrt{k}}.$$

По правилу трех стандартов доверительный интервал определяется по соотношениям:

$$\bar{x} - 3S_{cp} < \bar{x} < \bar{x} + 3S_{cp} \text{ для } \beta = 0,997;$$

$$\bar{x} - 2S_{cp} < \bar{x} < \bar{x} + 2S_{cp} \text{ для } \beta = 0,950;$$

$$\bar{x} - S_{cp} < \bar{x} < \bar{x} + S_{cp} \text{ для } \beta = 0,680;$$

Если выборка содержит меньше 30 значения, то предполагается, что случайная величина подчиняется распределению Стьюдента, а доверительный интервал рассчитывают по соотношению:

$$\bar{x} - t_{p,f} \frac{S}{\sqrt{k}} < \bar{x} < \bar{x} + t_{p,f} \frac{S}{\sqrt{k}},$$

где $t_{p,f}$ – коэффициент Стьюдента, который зависит от уровня значимости ($p = 1 - \beta$) и числа степеней свободы ($f = k - 1$).

Статическую оценку выборки средствами электронных таблиц можно осуществить с помощью сервиса «Описательная стати-

стика». Набор рассчитываемых параметров для выборки практически совпадает в Excel и Calc (рис. 1).

Excel		Calc	
Среднее	0,0554	Среднее	0,0554
Стандартная ошибка	0,0013	Стандартная ошибка	0,0013
Медиана	0,054	Мода	0,053
Мода	0,053	Медиана	0,054
Стандартное отклонение	0,0029	Первый квартиль	0,053
Дисперсия выборки	8E-06	Третий квартиль	0,058
Экспесс	-2,8495	Дисперсия	8E-06
Асимметричность	0,5897	Среднеквадратическое Отклонение	0,0029
Интервал	0,006	Экспесс	-2,8495
Минимум	0,053	Асимметрия	0,5897
Максимум	0,059	Диапазон	0,006
Сумма	0,277	Минимум	0,053
Счет	5	Максимум	0,059
Уровень надежности(95,0%)	0,0036	Сумма	0,277
		Количество	5

Рис. 1. Результаты применения сервиса «Описательная статистика» к одним и тем же данным в Excel и Calc.

Корреляция

При установлении взаимосвязи между двумя величинами исследователь, пытаясь визуализировать результат, наносит результаты в виде точек на диаграмму. Довольно часто полученный результат сразу невозможно описать линейной или др. зависимостью, т.к. на диаграмме точки занимают некоторую область плоскости (рис. 2в). Такие диаграммы называют диаграммами рассеивания (рис. 2). Причиной рассеивания являются неучтенные факторы, погрешности эксперимента и т.д.

Используя диаграмму рассеивания можно лишь с некоторой вероятностью оценить наличие взаимосвязи между откликом (y_i) и задаваемым параметром (x_i). С помощью диаграммы рассеивания можно лишь качественно оценить наличие корреляции между y_i и x_i .