

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ СРЕДСТВАМИ EXCEL В УЧЕБНОМ ПРОЦЕССЕ

In this paper we discuss the problem of teaching the statistical methods for students of the engineering specialities. We consider the basic statistical procedures for the experimental data processing. We note the efficiency of the use the Microsoft Excel functions for educational tasks on statistics.

Необходимость преподавания статистических методов студентам химико-технологических и других инженерных специальностей очевидна. Современные экспериментально-статистические методы имеют огромное методологическое значение для любого инженера-исследователя, поскольку позволяют ускорить изучение сложных объектов и значительно повысить его эффективность. По выражению специалиста в области управления качеством Исикава Кафу, «инженеры, которые дают оценку на основе экспериментальных данных, должны знать статистические методы наизусть». Управление качеством – это прежде всего применение современных статистических методов. В японских корпорациях все, начиная от председателя совета директоров и до рядового рабочего в цехе, обязаны знать хотя бы основы статистических методов. В США и Японии школьники учат статистические методы как один из обязательных школьных предметов. ЮНЕСКО регулярно проводит конференции по преподаванию статистики в средней школе. К сожалению, у нас ситуация совершенно иная. В средней школе методы статистического анализа данных (хотя многие из них очень просты и весьма полезны) не упоминаются вовсе, а в вузах прикладной статистике отводится очень скромное место.

При обработке любых экспериментальных данных необходимо грамотное применение статистических процедур. Важно понимать, в каких ситуациях применимы различные статистические методы, знать, каковы их свойства, уметь интерпретировать результаты. К сожалению, в настоящее время студентам многих специальностей статистические методы в рамках тех или иных курсов преподают специальные кафедры, не имеющие возможности уделить достаточное внимание теоретическому обоснованию рассматриваемых процедур и дать студентам понимание границ применимости этих методов.

В современных условиях, когда умение использовать компьютерные технологии стало неотъемлемой частью квалификации специалиста, преподавание статистики должно также вестись с использованием компьютера. Для эффективной подготовки специалистов используемые технологии и информационная среда

специалиста-технолога и студента должны быть идентичны. Поэтому, на наш взгляд, целесообразно использовать на занятиях в вузе достаточно простое, но удобное и широко распространенное программное средство – электронные таблицы Excel, которые теперь устанавливаются вместе с системой Windows на каждый компьютер. Пакет Microsoft Excel является наиболее простым и вместе с тем весьма эффективным средством, позволяющим проводить все основные процедуры статистической обработки экспериментальных данных. Простота использования позволяет сконцентрировать внимание студентов на статистическом содержании задач. С другой стороны, доступность пакета гарантирует его востребованность студентами сначала при написании курсовых и дипломных работ, а затем в профессиональной деятельности.

Рассмотрим основные статистические задачи, возникающие в практике химика-экспериментатора и инженера-исследователя, и инструменты пакета Excel, которые могут быть использованы при их решении.

Прежде всего назовем ряд предусмотренных в пакете Excel статистических функций: функции СРЗНАЧ и ДИСП вычисляют выборочное среднее и несмещенную оценку дисперсии; функция КОРРЕЛ вычисляет коэффициент корреляции между двумя массивами значений; функции НОРМСТОБР, ФРАСПОБР, СТЬЮДРАСПОБР, ХИ2ОБР возвращают квантили соответствующих распределений: стандартного нормального, Фишера, Стьюдента, χ^2 -распределения. С помощью этих функций могут быть решены задачи вычисления основных числовых характеристик выборки, построения доверительных интервалов для тех или иных показателей, оценки тесноты линейной связи между двумя наблюдаемыми показателями, проверки статистических гипотез, сравнения точности измерений различными приборами, сравнения средних значений некоторого показателя при двух различных способах действий и т. д. При этом важно помнить, что большинство широко распространенных статистических методов разработаны в предположении, что исследуемая выборка взята из нормального распределения [1, 2]. Именно в таких задачах оправдано использование связанных

с нормальным распределением Фишера, Стьюдента, χ^2 -распределением. Поэтому применению этих процедур, как правило, должно предшествовать обоснование нормальности распределения на основании проверки соответствующей статистической гипотезы либо на основании имеющейся априорной информации. Например, при исследовании мощности выпущенных на заводе электрических лампочек мы можем предположить, что эта мощность описывается нормальным распределением, а время службы электрической лампочки подчиняется экспоненциальному закону распределения. Статистическая проверка соответствия между выборочными данными и некоторым законом распределения осуществляется с помощью критериев, которые называются критериями согласия. Наиболее распространенным из них является критерий χ^2 Пирсона, основанный на сравнении эмпирических и теоретических частот.

Особое внимание следует обращать на наличие в выборках выбросов ошибочных, сильно отличающихся от основной массы наблюдений, поскольку даже одно такое грубое наблюдение способно сильно исказить выборочные характеристики и привести ко многим другим ошибочным выводам. Для обнаружения выбросов удобно использовать графическое представление данных с помощью точечной диаграммы или гистограммы. В случае выборки из нормального распределения проверка однородности результатов измерений может быть проведена с помощью критерия Стьюдента: если в выборке объема n одно наблюдение y^* сильно отличается от всех остальных, то с доверительной вероятностью α его следует признать ошибочным, если

$$t_{\text{расч}} = \frac{|y^* - \bar{y}|}{s\{y\}} > t_{\text{табл}} = t_{\alpha, n-2}, \quad (1)$$

где \bar{y} и $s^2\{y\}$ – выборочные среднее и дисперсия, рассчитанные по выборке, из которой исключено подозрительное наблюдение y^* ; $s\{y\} = \sqrt{s^2\{y\}}$; $t_{\alpha, n-2}$ – квантиль уровня α распределения Стьюдента с числом степеней свободы $\nu = n - 2$, который может быть найден по таблице или при помощи функции СТЬЮДРАСПОБР пакета Excel.

Критерии, основанные на распределении Стьюдента, используются также при проверке гипотез о средних значениях нормальных выборок в случае, когда дисперсии неизвестны и оцениваются по этим выборкам. Если дисперсии известны, то аналогичные статистики имеют стандартное нормальное распределение. Задачи сравнения средних могут возникать при оценке

соответствия некоторого показателя качества номинальному значению, при оценке эффективности новой технологии производства, при сравнении двух методов обработки, результатов действия двух лекарств и т. д. Однако процедура сравнения средних значений двух выборок зависит от того, однородны ли выборочные дисперсии, поэтому анализу средних должна предшествовать проверка гипотезы о равенстве дисперсий. Для задач сравнения нескольких (более двух) выборочных средних следует использовать методы дисперсионного анализа, предназначенного для исследования влияния тех или иных факторов на результаты эксперимента.

Задачи анализа выборочных дисперсий возникают при сравнении точности двух приборов или двух методов измерения. При оценке воспроизводимости эксперимента также необходимо проверить однородность дисперсий исследуемого параметра в различных опытах. Для выборок из нормального распределения проверка гипотезы о равенстве дисперсии заданному значению проводится с помощью критерия χ^2 , сравнение двух дисперсий осуществляется по критерию Фишера, нескольких – по критерию Бартлетта, который также приводит к распределению χ^2 . В случае нескольких выборок одинакового объема для проверки однородности дисперсий можно применять более удобный критерий Кохрена, основанный на отношении максимальной выборочной дисперсии к сумме всех дисперсий. Критерий Фишера также может быть использован для сравнения нескольких дисперсий: при этом проверяют гипотезу о равенстве максимальной и минимальной дисперсий.

Однако эти критерии могут применяться только при исследовании нормальных выборок, они не могут быть рекомендованы в случае отклонения от нормальности. Отметим, что при применении критерия Стьюдента к большим выборкам, несколько отличающимся от нормальных, результаты будут близки к правильным, если коэффициенты асимметрии и эксцесса такие же, как у нормального закона.

При отклонениях от нормальности для анализа однородности двух выборок следует использовать методы, свободные от распределения. Наиболее общепотребительными являются критерий Манна – Уитни, ранговый критерий Уилкоксона, а также критерии, основанные на знаках.

Многие практические задачи сводятся к определению функциональной зависимости некоторого параметра y , который называется откликом, от одного или нескольких независимых факторов. В случае одного фактора о виде уравнения зависимости можно судить по графику. Иногда примерный вид зависимости бывает известен из теоретических соображений или предыдущих исследований аналогичных

данных. Подбор аппроксимирующей функции в Excel может быть выполнен путем построения точечной диаграммы и добавления к ней линии тренда. В пакете Excel предусмотрено шесть типов линий тренда, которые могут быть добавлены к диаграмме: линейная, логарифмическая, полиномиальная с выбором степени полинома, степенная, экспоненциальная и линейная фильтрация (скользящее среднее). Определение коэффициентов выбранной модели производится путем линеаризации и применения метода наименьших квадратов. Вместе с уравнением зависимости на диаграмме может быть показан коэффициент детерминации R , позволяющий количественно оценить достоверность приближения. В случае линейной однофакторной зависимости коэффициент детерминации равен квадрату коэффициента корреляции.

В случае зависимости отклика от нескольких факторов x_1, x_2, \dots, x_k для определения коэффициентов и анализа множественного уравнения регрессии

$$\bar{y} = b_0 + b_1 x_1 + \dots + b_k x_k \quad (2)$$

в пакете Excel удобно использовать функцию ЛИНЕЙН, которая также основана на методе наименьших квадратов. Вывод результата действия функции ЛИНЕЙН осуществляется в следующем формате:

b_k	b_{k-1}	...	b_1	b_0
$s\{b_k\}$	$s\{b_{k-1}\}$...	$s\{b_1\}$	$s\{b_0\}$
R	$\sqrt{s_{ал}^2}$			
$F_{расч}$	$f_{ад}$			
$SS_{регp}$	$SS_{ост}$			

В первой строке записываются коэффициенты уравнения регрессии (2) в обратном порядке, во второй – их среднеквадратические отклонения, характеризующие точность оценок коэффициентов; первый элемент третьей строки – множественный коэффициент детерминации, остальные величины используются для статистического анализа уравнения регрессии. Отметим, что таким образом могут быть найдены и более сложные, нелинейные

по факторам, зависимости. Необходимо только, чтобы параметры b_0, b_1, \dots, b_k входили в уравнение линейно, т. е. чтобы при некоторой замене переменных уравнение регрессии сводилось к (2).

Перечисленные задачи, а также методы планирования и обработки результатов активного эксперимента при построении регрессионных моделей и поиске оптимальных условий рассматриваются автором в курсе «Планирование и организация эксперимента», который читается для студентов специальности «Физико-химические методы и приборы контроля качества продукции» в 4-м семестре [3]. Практические занятия по этому курсу проводятся в компьютерном классе. Студенты используют электронные таблицы Excel для выполнения индивидуальных заданий по анализу средних и дисперсий, подбору зависимостей, обработке результатов пассивного и активного эксперимента. Кроме того, составлена программа, моделирующая зависимость отклика от нескольких факторов, наблюдаемую на фоне случайных помех. С помощью этой программы студенты могут самостоятельно планировать и проводить активный эксперимент, задавая условия опытов. Выполнение расчетных заданий с применением пакета Excel дает студентам полезный опыт использования вычислительной техники для решения прикладных задач.

Литература

1. Джонсон, Н. Статистика и планирование эксперимента в технике и науке. Методы обработки данных / Н. Джонсон, Ф. Лион. – М.: Мир, 1980. – 610 с.
2. Тюрин, Ю. Н. Статистический анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров. – М.: ИНФРА-М, 1998. – 528 с.
3. Блинова, Е. И. Планирование и организация эксперимента: практикум по выполнению расчетных заданий с использованием ЭВМ для студентов II курса специальности 1-54 01 03 «Физико-химические методы и приборы контроля качества продукции» / Е. И. Блинова. – Минск: БГТУ, 2004. – 90 с.