

**ИСПОЛЬЗОВАНИЕ РЕГРЕССИОННОГО АНАЛИЗА ПРИ ОБРАБОТКЕ РЕЗУЛЬТАТОВ
ЭКСПЕРИМЕНТА ПО ОПРЕДЕЛЕНИЮ УГЛА ПОВОРОТА ПЛОСКОСТИ ПОЛЯРИЗАЦИИ
СВЕТА САХАРОЗОЙ**

Егоров Всеволод Михайлович

ученик

МБОУ «Гимназия № 3», г. Астрахань

Abstract: the results of the experiment aimed at determining the rotation angle of plane of light polarization by sucrose were processed employing regression analysis. The specific rotation of sucrose was defined, the selective regression equation was derived, the selective regression line was plotted, linearity deviations, the correlation coefficient of sucrose specific rotation and rotation angle of plane of light polarization by sucrose were calculated. The article presents the mathematical foundation of a device under development for determining quality of honey.

Key words: regression analysis, linear regression, least square method, selective regression equation, selective regression line, correlation coefficient, electro-optic activity, rotation angle of plane of light polarization, specific rotation, sucrose, device for determining quality of honey.

Одной из важнейших проблем в современном мире является определение качества пищевых продуктов. В ряду ценнейших для здоровья человека продуктов важное место занимает мед [7]. В то же время известно, что доля фальсифицированного мёда на рынке весьма высока [2]. Качество меда определяется высоким содержанием фруктозы и глюкозы и невысоким содержанием других веществ, в частности сахарозы, ее содержание в натуральном меде не должно превышать 4%.

Как показывает статистика, которую приводит Министерство предпринимательства Новой Зеландии, среднестатистическое годовое потребление мёда в России составляет 0,35 кг [8]. Это говорит о том, что есть спрос на данный продукт и, как следствие, растет предложение по контрафактному меду. Учитывая большие затраты при получении натурального мёда, количество контрафактного в последнее время заметно увеличилось.

Известно, что сахараиды, из которых почти полностью состоит мед, обладают электрооптической активностью, т. е. способностью вращать плоскость поляризации проходящего через слой вещества света.

Любые способы изготовления фальсифицированного мёда приводят к увеличению содержания в нём сахарозы, в связи с чем становится актуальной проблема разработки метода определения превышения концентрации содержащейся в мёде сахарозы выше нормы для натурального мёда.

Таким образом, был проведён эксперимент на установке для изучения закона Малюса производства Phywe Systems (EXACTA OPTECH) с целью определения зависимости удельного вращения чистой сахарозы в растворах от длины волны света. Установка, которая использовалась для этого, была модифицирована: монохроматический источник (натриевая лампа) был заменён на немонхроматический (лампу с широким спектром). При этом, в эксперименте применялся зелёный светофильтр с длиной волны $\lambda=530$ нм, который ранее в данной установке не использовался [3]. С помощью весов взвешивалась чистая сахароза для приготовления растворов.

При обработке экспериментальных данных для расчета различных параметров часто используют линейные соотношения между измеряемыми величинами или их функциями, откладывая попарно значения измеренных величин или их функций и проводя через полученные точки прямую. По отрезкам, отсекаемым этой прямой на осях координат, и тангенсу угла наклона прямой рассчитывают значения физических параметров. Для того чтобы провести прямую через множество точек с помощью ЭВМ наилучшим образом, необходимо сначала выяснить, каким критериям должна удовлетворять эта прямая [4]. В большинстве случаев прямую стараются провести так, чтобы сумма квадратов отклонений экспериментальных значений величины y от прямой была минимальна [5].

Для нахождения выборочного уравнения регрессии и построения по нему выборочной линии регрессии, а также вычисления значения удельного вращения согласно уравнению, проведем регрессионный анализ.

Выбрав вид функции регрессии, т.е. вид рассматриваемой модели зависимости Y от X (или X от Y), например, линейную модель $y_x=a+bx$, необходимо определить конкретные значения коэффициентов модели [6].

При различных значениях a и b можно построить бесконечное число зависимостей вида $y_x=a+bx$, т.е. на координатной плоскости имеется бесконечное количество прямых, нам же необходима такая зависимость, которая соответствует наблюдаемым значениям наилучшим образом. Таким образом, задача сводится к подбору наилучших коэффициентов [9].

Линейную функцию $a+bx$ ищем, исходя лишь из некоторого количества имеющихся наблюдений. Для нахождения функции с наилучшим соответствием наблюдаемым значениям используем метод наименьших квадратов [1].

Введем обозначения:

θ_{α_i} – удельное вращение сахарозы, вычисленное по уравнению $\theta_{\alpha_i} = a+b\Psi_i$.

Ψ_i и c_i - измеренные значения, где Ψ_i – угол поворота плоскости поляризации света сахарозой, c – концентрация сахарозы.

$\varepsilon_i = \Psi_i - \theta_{\alpha_i}$ - разность между измеренными и вычисленными по уравнению значениями.

В методе наименьших квадратов требуется, чтобы ε_i , разность между измеренными Ψ_i и вычисленными по уравнению значениям θ_{0_i} , была минимальной. Следовательно, находим коэффициенты a и b так, чтобы сумма квадратов отклонений наблюдаемых значений от значений на прямой линии регрессии оказалась наименьшей:

$$S = \sum_{i=1}^n [\varepsilon_i]^2 = \sum_{i=1}^n (\Psi_i - \theta_{0_i})^2 = \sum_{i=1}^n (\Psi_i - a - b\Psi_i)^2 \rightarrow \min \quad (1)$$

Исследуя на экстремум эту функцию аргументов a и b с помощью производных, можно доказать, что функция принимает минимальное значение, если коэффициенты a и b являются решениями системы:

$$\begin{cases} a \sum_{i=1}^n \Psi_i + b \sum_{i=1}^n \Psi_i^2 = \sum_{i=1}^n c_i \Psi_i \\ na + b \sum_{i=1}^n \Psi_i = \sum_{i=1}^n c_i \end{cases} \quad (2)$$

Если разделить обе части нормальных уравнений на n , то получим:

$$\begin{cases} a \frac{\sum_{i=1}^n \Psi_i}{n} + b \frac{\sum_{i=1}^n \Psi_i^2}{n} = \frac{\sum_{i=1}^n c_i \Psi_i}{n} \\ a + b \frac{\sum_{i=1}^n \Psi_i}{n} = \frac{\sum_{i=1}^n c_i}{n} \end{cases} \quad (3)$$

Учитывая, что $\frac{\sum_{i=1}^n \Psi_i}{n} = \bar{\Psi}$, $\frac{\sum_{i=1}^n c_i}{n} = \bar{c}$, $\frac{\sum_{i=1}^n \Psi_i^2}{n} = \overline{\Psi^2}$, $\frac{\sum_{i=1}^n c_i \Psi_i}{n} = \overline{c\Psi}$

Получим, $\begin{cases} a\bar{\Psi} + b\bar{\Psi} = \overline{c\Psi} \\ a + b\bar{\Psi} = \bar{c} \end{cases}$ отсюда $a = \bar{c} - b\bar{\Psi}$, подставляя значение a в первое уравнение, получим:

$$\bar{\Psi}(\bar{c} - b\bar{\Psi}) + b\bar{\Psi}^2 = \overline{c\Psi} \rightarrow \bar{c}\bar{\Psi} - b\bar{\Psi}^2 + b\bar{\Psi}^2 = \overline{c\Psi} \rightarrow b(\bar{\Psi}^2 - \overline{\Psi^2}) = \overline{c\Psi} - \bar{c}\bar{\Psi} \quad (4)$$

$$b = \frac{\overline{c\Psi} - \bar{c}\bar{\Psi}}{\bar{\Psi}^2 - \overline{\Psi^2}} \quad (5)$$

$$a = \frac{\bar{\Psi}^2 \bar{c} - \overline{\Psi c\Psi}}{\bar{\Psi}^2 - \overline{\Psi^2}} \quad (6)$$

При этом b называют коэффициентом регрессии; a называют свободным членом уравнения регрессии и вычисляют по формуле: $a = \bar{c} - b\bar{\Psi}$.

Полученная прямая является оценкой для теоретической линии регрессии. Имеем:

$$\theta_0 = a + b\Psi = \bar{c} - b\bar{\Psi} + b\Psi = \bar{c} + b(\Psi - \bar{\Psi}) \quad (7)$$

Итак, $\theta_0 = \bar{c} + b(\Psi - \bar{\Psi})$ является уравнением линейной регрессии.

Таким образом у нас имеется 10 экспериментальных значений угла поворота плоскости поляризации света и концентрации [Табл. 1].

Таблица 1.

c_i	2	5	7	9	10	3	6	4	1	8
Ψ_i	3	6	8	11	12	4	7	5	2	9

Чтобы найти выборочное уравнение регрессии θ_0 на, сначала проведем упорядочивание данных по значениям Ψ_i и c_i . Получаем новую таблицу [Табл. 2]:

Таблица 2.

c_i	1	2	3	4	5	6	7	8	9	10
Ψ_i	2	3	4	5	6	7	8	9	11	12

Для упрощения вычислений составим расчетную таблицу, в которую занесем необходимые численные значения [Табл. 3].

Таблица 3.

c_i	Ψ_i	Ψ_i^2	$\Psi_i c_i$
1	2	4	2
2	3	9	6
3	4	16	12
4	5	25	20
5	6	36	30
6	7	49	42
7	8	64	56
8	9	81	72
9	11	121	99
10	12	144	120
$\sum c_i = 55$	$\sum \Psi_i = 67$	$\sum \Psi_i^2 = 549$	$\sum \Psi_i c_i = 459$
$c = 5,5$	$\Psi = 6,7$	$\Psi^2 = 54,9$	$\Psi c = 45,9$

Согласно формуле (5), вычисляем значение коэффициента регрессии

$$b = \frac{45,9 - 5,5 * 6,7}{54,9 - 6,7} = \frac{45,9 - 36,85}{48,2} = \frac{9,05}{48,2} = 0,1876$$

а по формуле (6)

$$a = \frac{54,9 * 5,5 - 6,7 * 45,9}{54,9 - 53,5} = \frac{301,95 - 307,53}{1,4} = -3,986$$

Таким образом, выборочное уравнение регрессии имеет вид $y = -3,986 + 0,1876x$.

Нанесем на координатной плоскости точки и отметим прямую регрессии (график 1).

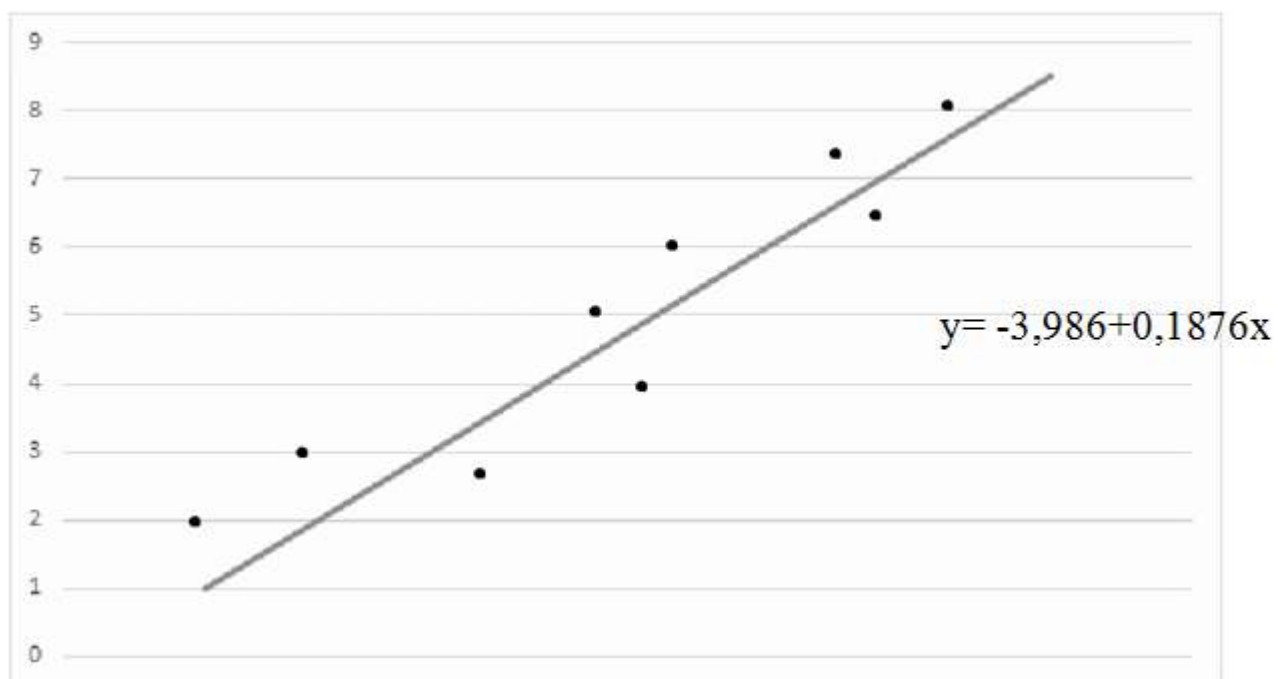


График 1.

На графике видно, как располагаются наблюдаемые значения относительно линии регрессии. Для численной оценки отклонений от линейности, составим таблицу [Табл. 4]:

Таблица 4.

Ψ_i	c_i	θ_{σ_i}	$\theta_{\sigma_i} - \Psi_i$
2	1	1,505	0,505
3	2	1,869	-0,131
4	3	3,584	0,584
5	4	4,463	0,463
6	5	5,734	0,734
7	6	6,648	0,648
8	7	6,828	-0,172
9	8	8,585	0,585
11	9	10,148	1,148
12	10	11,342	1,342

Значения θ_{σ_i} вычислены согласно уравнению регрессии.

Заметное отклонение некоторых наблюдаемых значений от линии регрессии объясняется малым числом наблюдений. При исследовании степени линейной зависимости θ_{σ} от Ψ число наблюдений учитывается. Сила зависимости определяется величиной коэффициента корреляции.

Коэффициент корреляции - величина, которая может варьировать в пределах от +1 до -1. В случае полной положительной корреляции этот коэффициент равен +1, а при полной отрицательной – -1. В нашем случае коэффициент корреляции, высчитанный в программе Excel, составил 0,995874, что говорит о почти полной положительной корреляции. Коэффициент корреляции, высчитанный в программе Excel, составил 0,995874, что говорит о почти полной положительной корреляции.

Результаты проведенного математического анализа и полученная математическая модель могут быть использованы при создании малогабаритного малоинерционного анализатора для определения качества меда.

Список литературы:

1. Благовещенский Ю.В. Вычисление элементарных функций на ЭВМ / - М.: Наука - 240с.
2. Бобылёв С. Доля фальсификата на рынке меда в России / ТАСС - информационное агентство России [Электронный ресурс] - URL: <http://tass.ru/ekonomika/3542884> (дата обращения: 07.01.18)
3. Егоров В.М. Куриков И.А. Лихтер А.М. С.З. Субботина. Разработка метода идентификации фальсифицированного мёда / Norwegian Journal of Development of the International Scenes. №15. -Изд. 1. -Осло: 2017. - С. 55-58.
4. Ландау Л.Д. Лившиц Е.М. Теоретическая физика. – В 10 т. / Учебное пособие для вузов. Т. I. Механика. - 5-е изд., стереот. - М: ФИЗМАТЛИТ, 2004. - 224 с.
5. Математическая энциклопедия. / Ред. И.М. Виноградова. -Т.2. М.: Советская Энциклопедия, 1979. – 552 с.
6. Немыцкий В.В. М.И. Слудская. ЧСеркасов А.Н. Курс математического анализа / –Т. 2. - М.: Гостехиздат, 1944. – 405 с.
7. Николаева Ю.Н., Мед, прополис, перга и другие продукты пчеловодства от всех болезней / - М.: Рипол Классик, 2011. - 192 с.
8. Пономарёв А. Мировой рынок меда в 2017 году / Мир пчеловодства [Электронный ресурс] - URL: <http://www.apeworld.ru/1460449678.html> (дата обращения: 17.03.18)
9. Шашков В.Б. Прикладной регрессионный анализ. / –Оренбург: Издательский дом ГОУ ВПО ОГУ, 2003. – 363 с.