

3. Икрами М.Б., Мирзорахимов К.К., Шарипова М. Способы окрашивания пищевых продуктов натуральными пищевыми красителями // Труды ТУТ. – 2007. – Т. 12. – С. 106-110.

ОПТИМИЗАЦИЯ ИЗМЕРЕНИЯ РАССТОЯНИЙ ПРИ ИСПОЛЬЗОВАНИИ МОДЕЛИ ИЗОЛЯЦИИ РАССТОЯНИЕМ МАЛЕКО ДЛЯ АНАЛИЗА ПОПУЛЯЦИОННО-ГЕНЕТИЧЕСКИХ ПАРАМЕТРОВ¹

© Симонов Ю.И.^{*}, Ельчинова Г.И.[♦], Зинченко Р.А.[♥]

Медико-генетический научный центр РАМН, г. Москва

В данной статье освещен алгоритм, оптимизирующий измерения расстояний при использовании «модели изоляции расстоянием Малек» для анализа популяционно-генетических параметров. Алгоритм реализован стандартными функциями программы «Excel». Возвращает расстояния между 2-мя пунктами, используя их GPS координаты.

Использование небиологических источников информации является традиционным при популяционно-генетических исследованиях и позволяет получать оценки ряда параметров для больших территорий. На основании брачных записей мы можем оценить этническую ассортативность, эндогамность и параметры изоляции расстоянием Малек [3], основным из которых является локальный инбридинг a [1]. Суть этого метода заключается в установлении зависимости степени кровного родства супругов от расстояния между местами их рождения. Модель популяционной структуры, построенная на представлении об изоляции расстоянием Малек, предполагает, что население равномерно распределено по изучаемой территории, брачные миграции изотропны (т.е. происходят отовсюду достаточно равномерно), а вероятность каждой из них обратно пропорциональна расстоянию между местами рождения супругов. Сосредоточение населения в отдельных населенных пунктах с выраженной дисперсией их размеров ограничивает адекватность этой модели, однако, этим обстоятельством можно пренебречь, если изучаемая территория и, соответственно, количество поселений на ней достаточно велико. Расстоя-

¹ Работа выполнена при частичном финансировании РФФИ (10-04-00411, 11-04-00012, 12-06-00004) и ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технического комплекса России на 2007-2012 годы».

^{*} Лаборант-исследователь лаборатории Генетической эпидемиологии МГНЦ РАМН, студент МАИ 5-го курса.

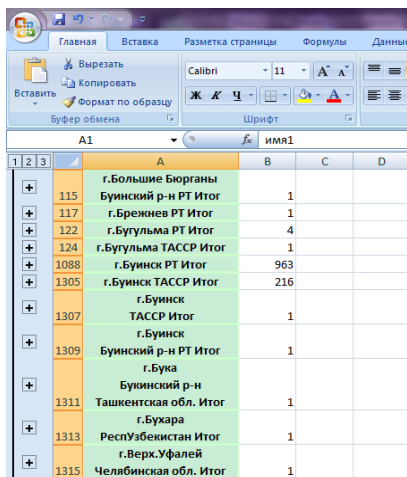
[♦] Ведущий научный сотрудник лаборатории Генетической эпидемиологии МГНЦ РАМН, доктор биологических наук.

[♥] Руководитель лаборатории Генетической эпидемиологии МГНЦ РАМН, доктор медицинских наук, профессор.

ния между местами рождения супругов измеряются по прямой в километрах. При отсутствии точной информации, расстояния можно измерять с точностью до районов или областных центров. Однако часть информации тем или иным образом оказывается утерянной или искаженной [2]. Несмотря на кажущуюся простоту метода именно «ручное» измерение расстояний отпугивает исследователей от использования модели изоляции расстоянием, которая часто оказывается единственным методом для получения популяционно-генетических характеристик. Задачей настоящей работы была оптимизация измерения расстояний при использовании модели изоляции расстоянием Малеко для анализа популяционно-генетических параметров. В качестве модельного был выбран Буинский район РТ. Население района превышает 45 тыс. человек. Брачные записи выкопированы за 1990-2000 гг. Из рассмотрения, как и ранее, исключены браки супругов пострепродуктивного возраста (женщина старше 45 лет) и браки жителей других регионов, приезжающих играть свадьбу на историческую родину. В анализ вошли 3422 брачные записи. Каждая брачная запись имеет сведения о возрасте брачующихся, их национальности, месте рождения и месте жительства, представленные в виде таблицы (Excel). Традиционно замеры расстояний производятся «вручную» посредством инструментов, предлагаемых интернет сервисом «Яндекс карты» (<http://maps.yandex.ru>) Для этого требуется выполнить следующие манипуляции:

1. Ввести в поисковую строку название первого населенного пункта.
2. Поставить маркер линейки.
3. Ввести в поисковую строку название второго населенного пункта.
4. Поставить маркер линейки.
5. Ввести полученное значение в соответствующее поле таблицы.

Описанная процедура осложнялась следующими моментами. Во-первых, исследования проводились на достаточно ограниченном пространстве, многие название населенных пунктов, в процессе измерения часто повторялись. Например, при анализе Буинского района, кол-во людей из г. Буинска, участвовавших в исследовании, достигало 1179 человек (рис. 1). Т.е. получается, что было необходимо повторять первые 2 пункта из описанного алгоритма более 1000 раз, Кроме того существует множество комбинаций из одних и тех же названий, что добавляет количество ненужных манипуляций. Современные компьютерные технологии позволяют избежать лишних действий, освобождая тем самым научный персонал от механической работы. Предлагается следующая схема оптимизации измерения расстояний при использовании модели изоляции расстоянием Малеко: Создается справочник оригинальных названий населенных пунктов. Эта операция осуществляется элементарно, посредством стандартных фильтров Excel. Т.о. в этот справочник г. Буинск заносится один раз. Далее, посредством любого из географических сервисов, этому населенному пункту присваиваются его GPS координаты. Когда координаты всех населенных пунктов (один раз для каждого) введены, происходит автоматический расчет расстояния, по координатам.



	A	B	C	D
115	г. Большие Бюрганы Буинский р-н РТ Итог	1		
117	г. Брежнев РТ Итог	1		
122	г. Бугульма РТ Итог	4		
124	г. Бугульма ТАССР Итог	1		
1088	г. Буинск РТ Итог	963		
1305	г. Буинск ТАССР Итог	216		
1307	г. Буинск ТАССР Итог	1		
1309	Буинский р-н РТ Итог	1		
1311	г. Бука Буинский р-н Ташкентская обл. Итог	1		
1313	г. Бухара Респ. Узбекистан Итог	1		
1315	г. Верх. Уфaley Челябинская обл. Итог	1		

Рис. 1. Повторяемость названий населенных пунктов

Алгоритм предлагаемого выше будет выглядеть следующим образом:

1. Однократное формирование справочника.
2. Ввод в поисковую строку названия населенного пункта.
3. Копирование координат в соответствующее поле.

Таким образом, получается всего 3 итерации, а самое главное – количество обращений к географическому сервису уменьшилось в несколько раз!

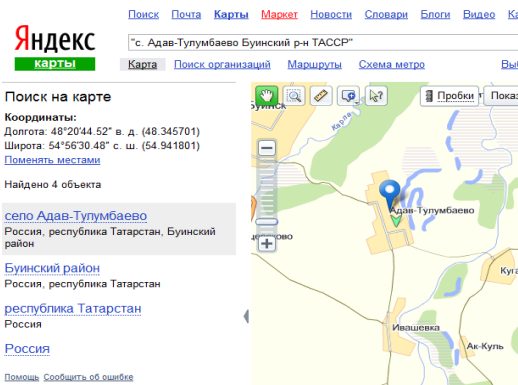


Рис. 2. Поиск населенного пункта в Яндекс-картах

На примере Буинского района будет проиллюстрировано выполнение алгоритма:

1. Однократное формирование справочника (рис. 1).
2. Ввод в поисковую строку названия населенного пункта (рис. 2).
3. Копирование координат в соответствующее поле (рис. 3).

	R	W	AB
1			
2			
3	Справочник ориг. названий	Долгота, Широта (ячейка для ввода)	
446	с. Адав-Тулумбаево Буйнский р-н РТ	48.341581, 54.944371	
447	с. Адав-Тулумбаево Буйнский р-н РТ	48.341581, 54.944371	
448	с. Адав-Тулумбаево Буйнский р-н РТ	48.341581, 54.944371	
449	с. Адав-Тулумбаево Буйнский р-н РТ	48.341581, 54.944371	
450	с. Адав-Тулумбаево Буйнский р-н ТАССР	48.341581, 54.944371	
451	с. Азатова Апастовский р-н РТ		

Рис. 3. Копирование координат в соответствующее поле.

В процессе построения алгоритма были выявлены следующие проблемы, не позволяющие обеспечить полностью автоматизированный процесс (такой, чтобы на основе имеющегося справочника, программа сама делала запрос на географический сервер и возвращала координаты):

1. Значительной части сел / поселков присвоены некорректные или ныне не актуальные районы и области.
2. В изначальных данных допущены грамматические ошибки при вводе. Автоматическое исправление которых могут осуществить только очень сложные и дорогостоящие программные решения.
3. Не стандартизировано заполненные формы. Адрес должен быть записан в строгом порядке.
4. Частые сокращения в обозначении места жительства.
5. Исчезнувшие к настоящему времени населенные пункты
6. Неправильно указаны координаты в интернет-данных (очень редко).

Для реализации описанной функции использовалась формула расчета по координатам, которая имеет следующий вид:

$$= (6372795 * (\text{ATAN}(\text{КОРЕНЬ}((\text{COS}((\text{ПИ}() * 1) * (\text{P}4058) / 180) * (\text{SIN}(((\text{ПИ}() * 1) * \text{Q}4058) / 180) - ((\text{ПИ}() * 1) * (\text{H}4058) / 180))))^2 + (\text{COS}((\text{ПИ}() * 1) * (\text{G}4058) / 180) * \text{SIN}((\text{ПИ}() * 1) * (\text{P}4058) / 180) - \text{SIN}((\text{ПИ}() * 1) * (\text{G}4058) / 180) * \text{COS}((\text{ПИ}() * 1) * (\text{P}4058) / 180) * \text{COS}(\text{ABS}(((\text{ПИ}() * 1) * \text{Q}4058) / 180) - ((\text{ПИ}() * 1) * (\text{H}4058) / 180))))^2) / (\text{SIN}((\text{ПИ}() * 1) * (\text{G}4058) / 180) * \text{SIN}((\text{ПИ}() * 1) * (\text{P}4058) / 180) + \text{COS}((\text{ПИ}() * 1) * (\text{G}4058) / 180) * \text{COS}((\text{ПИ}() * 1) * (\text{P}4058) / 180) * \text{COS}(\text{ABS}(((\text{ПИ}() * 1) * \text{Q}4058) / 180) - ((\text{ПИ}() * 1) * (\text{H}4058) / 180)))))) / 1000$$

Где P4058 ячейка в 4058й строке, P-го столбца. По аналогии Q4058, H4058, G4058, P4058. ATAN – арктангенс от числа. ПИ – возвращает число пи. Число 6372795 – средний радиус Земли. SIN, COS – функции синуса и косинуса, соответственно. При последующих сборах информации, необходимо использовать автоматизированное анкетирование (заполнение анкетных данных в специально разработанную форму и автоматизированное распознавание). После того, как расстояния вычислены, параметры изоляции расстоянием вычисляются стандартным способом, неоднократно приводимым нами в литературе. Для Буинского района $a = 0,00016$, $b = 0,0091$, $\sigma = 77,7$. При помощи сортировки можно, например, из базы данных отобрать только моноэтнические татарские браки, в этом случае $a = 0,00024$, $b = 0,0092$, $\sigma = 80,2$. Сразу становится заметно, что локальный инбридинг в татаро-татарских браках в 1,5 раза выше, чем в целом по Буинскому району. В дальнейшем при постановке новых задач перед исследователем сортировку можно осуществлять по годам, сельсоветам и пр. Экономическая эффективность предлагаемого алгоритма очевидна. Он дает возможность не выполнять рутинную работу научным персоналом, тем самым освобождая время на исследовательскую деятельность. Кроме того, введенные координаты можно использовать при картографическом анализе популяционно-генетической структуры, тем самым экономя рабочее время еще одного исследователя.

Список литературы:

1. Генетическая структура наследственные болезни чувашской популяции / Под ред. Е.К. Гинтера, Р.А. Зинченко. – Чебоксары: «Пегас». 2006. – С. 19-40.
2. Ельчинова Г.И. Методы обработки популяционно-генетических данных: структура брачных миграций // Медицинская генетика. – 2004. – Т. 3, № 5. – С. 185-192.
3. Morton N.E. Isolation by distance in human populations // Ann. Hum. Genet. – 1977. – № 40. – P. 361-365.

УФ-МУТАГЕНЕЗ, СВЕТОВАЯ И ТЕМНОВАЯ РЕПАРАЦИИ В ОПЫТАХ С ЛУКОМ-ПОРЕЕМ (*ALLIUM PORRUM*)

© Сироткина С.М.^{*}, Пак И.В.[♦], Дурасова О.Н.

Тюменский государственный университет, г. Тюмень

Изучена эффективность процессов световой и темновой репарации после воздействия ультрафиолетового облучения (УФО) в дозе 20 тыс. эрг в опытах с тест-объектом луком-пореем (*Allium porrum*). Показано, что УФО в изученной дозе снижает показатели выживаемости, измен-

^{*} Аспирант кафедры Экологии и генетики.

[♦] Профессор кафедры Экологии и генетики, доктор биологических наук.