Об определении объема бесповторной выборки при проведении медицинских исследований

Conference Paper · May 2017

2 authors:

Oleh Pihnastyi
Kharkiv Polytechnical Institute

Olga Kozina
Kharkiv National Medical University

НИУ БелГУ
BELGOROD STATE UNIVERSITY (BSU)

Сохраним лучшее!
Приумножим достигнутое!
Сделаем это вместе!

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»
ИНСТИТУТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ И ЕСТЕСТВЕННЫХ НАУК

ЕСТЕСТВЕННОНАУЧНЫЕ, ИНЖЕНЕРНЫЕ И ЭКОНОМИЧЕСКИЕ ИССЛЕДОВАНИЯ В ТЕХНИКЕ, ПРОМЫШЛЕННОСТИ, МЕДИЦИНЕ И СЕЛЬСКОМ ХОЗЯЙСТВЕ

МАТЕРИАЛЫ I МОЛОДЁЖНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

Белгород 2017

Министерство образования и науки РФ
Федеральное государственное автономное образовательное
учреждение высшего образования
«Белгородский государственный национальный
исследовательский университет»

ЕСТЕСТВЕННОНАУЧНЫЕ, ИНЖЕНЕРНЫЕ И ЭКОНОМИЧЕСКИЕ ИССЛЕДОВАНИЯ В ТЕХНИКЕ, ПРОМЫШЛЕННОСТИ, МЕДИЦИНЕ И СЕЛЬСКОМ ХОЗЯЙСТВЕ

Материалы I Молодёжной научно-практической конференции
с международным участием

Белгород 2017

УДК [62:61:63]:001.8
ББК 3В6+5В6+4В6
Е 86

Рекомендовано к изданию редакционно-издательским советом института инженерных технологий и естественных наук НИУ «БелГУ» (протокол № 6 от 12.04.2017).

Рецензенты:

Е.И. Евтушенко, доктор технических наук, профессор, зав. кафедрой технологии стекла и керамики, проректор по научной работе ФГБОУ ВО «Белгородский государственный технологический университет им. В.Г. Шухова»;

А.Н. Петин, доктор географических наук, профессор, декан факультета горного дела и природопользования НИУ «БелГУ»

Е 86 Естественнонаучные, инженерные и экономические исследования в технике, промышленности, медицине и сельском хозяйстве: материалы I Молодёжной научно-практической конференции с международным участием; под общ. ред. С.Н. Девицыной. – Белгород: ИД «Белгород» НИУ «БелГУ», 2017. − 693 с.

ISBN 978-5-9571-2298-2

Сборник включает статьи, посвященные широкому спектру актуальных проблем в области информационных технологий, техники, промышленности, медицины и сельского хозяйства. Статьи представляют следующие направления:

Информационные системы и технологии в технике, экономике, экологии и медицине;

Информационно-телекоммуникационные технологии и информационная безопасность;

Биотехнические системы и комплексы;

Биомедицинские технологии, медицинская физика;

Бизнес-информатика (е-бизнес, электронный бизнес);

Математическое, имитационное, ситуационное и 3D моделирование;

Системы SMART CITY;

Суперкомпьютеры и высокопроизводительные вычисления;

Решение математических задач в сфере науки, образования, управления и экономики;

Дифференциальные уравнения, динамические системы и оптимальное управление.

УДК [62:61:63]:001.8
ББК 3В6+5В6+4В6

ISBN 978-5-9571-2298-2

Определены основные характеристики биосенсора на основе печатного электрода, модифицированного смесью с выбранным соотношением компонентов: коэффициент чувствительности составил 1140±60 нА/мМ, нижняя граница – 0,237 мМ, предел обнаружения – 0,079 мМ. Определены лимитирующие стадии процессов, происходящих на модифицированном печатном электроде для выявления затруднений диффузии субстрата фермента в объеме золь-гель матрицы.

Дальнейшая работа по оптимизации процесса модификации печатного электрода должна быть направлена либо на уменьшение толщины слоя золь-гель матрицы, либо на изменение ее состава для снижения диффузионных затруднений внутри нее.

Работа выполнена при поддержке гранта РФФИ и Правительства Тульской области № 16-48-710959 р_а (договор ДС/44) и гранта Президента Российской Федерации для государственной поддержки молодых российских ученых – кандидатов наук, договор № 14.Z56.16.5425-МК.

Список использованных источников

1. Reshetilov A.N. Biosensors and biofuel cells: Research focused on practical application (Review) // Appl. Biochem. Microbiol. 2015. V. 51. I. 2. P. 264–269.

2. Lim J.W. et al. Review of Micro/Nanotechnologies for Microbial Biosensors // Front. Bioeng. Biotechnol. 2015. V. 3. I. P. 61.

3. O.N. Ponamoreva, O.A. Kamanina, V.A. Alferov, A.V. Machulin, T.V. Rogova, V.A. Arlyapov, et al., Yeast-based self-organized hybrid bio-silica sol-gels for the design of biosensors, Biosens. Bioelectron. 67 (2015) P. 321–326.

4. Albery W.J., Bartlett P.N. Amperometric enzyme electrodes. Part 1. Theory. J. Electroanal. Chem. 1985. V. 194. P. 211-222.

Секция ***БИОМЕДИЦИНСКИЕ ТЕХНОЛОГИИ, МЕДИЦИНСКАЯ ФИЗИКА***

ОБ ОПРЕДЕЛЕНИИ ОБЪЕМА БЕСПОВТОРНОЙ ВЫБОРКИ ПРИ ПРОВЕДЕНИИ МЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

Кожина О.С.
г. Харьков, «Харьковский Национальный медицинский университет»,
14.01.10 – «Педиатрия»
Пигнастый О.М.
г. Харьков, Национальный технический университет, «Харьковский политехнический институт»

Введение

Конечной целью выборочного наблюдения является характеристика признака Y генеральной совокупности на основе данных выборки объемом n [1]. Пусть изучается генеральная совокупность относительно количественного признака Y объемом N. Если значения признака Y соответственно $\{y_1, y_2 \ldots\ldots, y_{N-1}, y_N\}$, то генеральное среднее $\bar{y}$ есть среднее арифметическое значений признака генеральной совокупности [1,с.198]

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \ . \qquad (1)$$

Полагаем, что для изучения генеральной совокупности относительно количественного признака Y извлечена выборка объемом n. Пусть значения количественного признака X_w выборочной совокупности соответственно $\{x_{w,1}, x_{w,2} \ldots\ldots x_{w,(n-1)}, x_{w,n}\}$ [1,с.199]

$$x_w = \frac{1}{n}\sum_{i=1}^{n} x_{w,i} \ , \qquad w = 1.W \ . \qquad (2)$$

Выборочную среднюю x_w, найденную по данным одной выборки следует рассматривать как случайную величину X_w, а значит можно говорить о распределении выборочной средней и о числовых характеристиках этого распределения. Действительно, выборку возможно сделать различными способами, при которых количественный признак X_w выборочной совокупности приобретет значения $\{x_1, x_2 \ldots\ldots x_{(W-1)}, x_W\}$, где W – возможное количество вариантов осуществить выборку.

Пример. Пусть имеется множество генеральной совокупности Y (объем $N=3$) со значениями признака $\{y_1=0, y_2=1, y_3=0\}$. Если выборка бесповторная, то возможное число вариантов осуществить выборку

$$C_N^n = N!/((N-n)!\,n!). \tag{3}$$

При объеме $n=2$ имеем $C_3^2 = 3!/(1!\cdot 2!) = 3$: $\{x_{1,1}=y_1=0, x_{1,2}=y_2=1\}$, $x_1=0.5$, $\{x_{2,1}=y_1=0, x_{2,2}=y_3=0\}$, $x_2=0.0$, $\{x_{3,1}=y_2=1, x_{3,2}=y_3=0\}$, $x_3=0.5$. Таким образом, случайная величина X_w для 1-ой, 2-ой и 3-ей принимает значения в соответствие с выражением (2), равные $x_1=0.5$, $x_2=0.0$, $x_3=0.5$. Если способ выборки является повторным (после извлечения y_i его значение фиксируется и возвращается обратно с возможностью повторного изъятия), возможное количество комбинаций $n\cdot n$=9.

Значение количественного признака Y генеральной совокупности связано со значением количественного признака X_w выборочной совокупности через соотношение [1,с.217]:

$$P(|X_w-\bar{y}|<\Delta)=2\Phi\left(\frac{\Delta}{\sigma(X_w)}\right)=2\Phi(t), \qquad 2\Phi(t)=\gamma \tag{4}$$

$$\Delta = t\sigma(X_w), \tag{5}$$

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt, \tag{6}$$

где $\Phi(t)$– функция Лапласа, значение которой может быть получено из таблицы, например [2,с.473]. Выражение $|X_w-\bar{y}|<\Delta$ означает, что отклонение случайной величины X_w, определенной равенством (2) от среднего арифметического признака генеральной совокупности не превышает погрешность, равную значению Δ. Это неравенство может быть записано в следующем виде

$$x_w-\Delta<\bar{y}<x_w+\Delta. \tag{7}$$

Запись (7) означает: если известна выборочная средняя x_w (2), то можно утверждать, что с вероятностью γ истинное значение (1) количественного признака Y генеральной совокупности находит в пределах от $(x_w-\Delta)$ до $(x_w+\Delta)$. Величина $\sigma(X_w)$ - есть среднеквадратичное отклонение количественного признака X_w выборочной совокупности. Величина $\sigma(X_w)$, в общем случае, зависит от объема выборки n

$$\sigma(X_w)\approx f_\sigma(n). \tag{8}$$

1. Основные задачи при использовании выборочного метода

Систему уравнений (4) возможно представить в виде двух уравнений:

$$\frac{\Delta}{\sigma(X_w)}=t, \tag{9}$$

$$2\Phi(t)=\gamma \tag{10}$$

с четырьмя неизвестными: Δ, t, $\sigma(X_w)$, γ. Для ее разрешения требуется задать дополнительно два уравнения. В связи с этим при использовании выборочного метода возникают три основные задачи [2]:

Задача №1. Определить объем выборки n, необходимый для получения с требуемой точностью Δ результатов при заданной вероятности γ.

Полагается, что в результате выполненной выборки, при которой будет получена выборочная совокупность со значениями $\{x_{w,1}, x_{w,2}\ldots\ldots x_{w,(n-1)}, x_{w,n}\}$, определим выборочную среднюю x_w (2). Известна вероятность γ, с которой значение количественного признака Y генеральной совокупности $\bar{y}$ будет удовлетворять неравенству (7) с заданной погрешностью Δ. При этом следует иметь ввиду, если заданная вероятность равна $\gamma=0.95$, то в 50 случаем из 1000 равенство (7) выполняться не будет. Таким образом, из четырех неизвестных: Δ, t, $\sigma(X_w)$, γ в первой задаче задается Δ, γ (и соответственно t), а определяется $\sigma(X_w)$, из которого находим объем выборки n.

Задача №2. Определить возможный предел ошибки репрезентативности, гарантирующий результаты с заданной вероятностью и сравнить его с величиной допустимой погрешности.

Полагается что задан объем выборки n и вероятность γ, с которой будет выполнено соотношение (7). Требуется определить погрешностью Δ и сравнить ее с допустимой погрешностью для выполнения экспериментов. При постановки задачи заданными считаются вероятность γ (или t), объем выборки n (соответственно $\sigma(X_w)$), а требуется найти величину погрешности Δ.

Задача №3. Определить вероятность того, что ошибка выборки не привысит допустимой погрешности.

В данном случае задан объем выборки n и погрешность Δ. Требуется определить вероятность того, что ошибка выборки не привысит допустимой погрешности. При постановки задачи заданными считаются вероятность γ (или соответствующий ей параметр t), объем выборки n (соответственно $\sigma(X_w)$). Необходимо отыскать вероятность γ того, что ошибка выборки не привысит допустимой погрешности Δ.

2. Определение объема выборки n при бесповторной выборки

В случае бесповторного метода выбора элементов объема n, взятых из общей совокупности N для обследования, общее число возможных выборок определяется комбинаторной формулой (3). Будем полагать, что способ извлечения элементов объема n из генеральной совокупности N элементов таков, что каждая выборка из общего количества (3) имеет равную вероятность быть отобранной. Такой способ называется случайным отбором. Как и говорилось ранее, выбранный элемент не возвращается обратно в генеральную совокупность.

Выборочное среднее $\bar{x}_w$

$$\bar{x}_w = \frac{\sum_{w=1}^{C_N^n} x_w}{C_N^n} \qquad (11)$$

случайной величины X_w есть несмещенная оценка среднего значения $\bar{y}$ (1) для совокупности Y:

$$M[X_w] = \bar{x}_w = \frac{\sum_{w=1}^{C_N^n} x_w}{C_N^n} = \frac{\sum_{w=1}^{C_N^n}\sum_{i=1}^{n} x_{w,i}}{nC_N^n} = \frac{\sum_{w=1}^{C_N^n}\sum_{i=1}^{n} y_i}{n\,N!/\left((N-n)!\,n!\right)} = \frac{\sum_{i=1}^{N} y_i}{N} \qquad (12)$$

В силу того, что

$$M[Y] = \frac{1}{N}\sum_{i=1}^{N} y_i \ , \qquad (13)$$

следует

$$M[Y] = M[X_w] \, . \qquad (14)$$

Дисперсия случайной величины X_w определяется выражением

$$M\left[(X_w - \bar{y})^2\right] = \frac{\sum_{w=1}^{C_N^n}(x_w - \bar{y})^2}{C_N^n} \qquad (15)$$

Подставляя x_w в (15), запишем

$$\sigma^2(X_w) = M\left[(X_w - \bar{y})^2\right] = \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2 = \frac{N-n}{n(N-1)}D(Y) = \frac{N-n}{n(N-1)}\sigma^2(Y) \, , \qquad (16)$$

где

$$D(Y) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2 \ , \qquad (17)$$

Выражение (16) позволяет получить выражение для вычисления среднеквадратичного отклонения случайной величины X_w

$$\sigma^2(X_w) = \frac{N-n}{n(N-1)}\sigma^2(Y) \, . \qquad (18)$$

Используя (4–6), получим зависимость между погрешностью Δ, вероятностью γ и объемом выборки n

$$\Delta = t\sigma(X_w) = t\sigma(Y)\sqrt{\frac{N-n}{n(N-1)}} \, . \qquad (19)$$

Последнее равенство приведем к виду

$$\Delta^2 = t^2\sigma^2(Y)\frac{N-n}{n(N-1)} \, , \qquad (20)$$

$$\Delta^2 n(N-1) = t^2\sigma^2(Y)N - t^2\sigma^2(Y)n \, , \qquad (21)$$

$$n\cdot\left(\Delta^2(N-1) + t^2\sigma^2(Y)\right) = t^2\sigma^2(Y)N \, , \qquad (22)$$

позволяющему выразить объем выборки n через погрешностью Δ и параметр t, соответствующий вероятности γ

$$n = \frac{t^2\sigma^2(Y)N}{\Delta^2(N-1) + t^2\sigma^2(Y)}N \, , \qquad (23)$$

Если случайная величина Y имеет биномиальное распределение с математическим ожиданием

$$M[Y]=p \tag{24}$$

и дисперсией

$$D(Y)=\sigma^2(Y)=pq\,, \tag{25}$$

$$p+q=1\,, \tag{26}$$

то выражение (23) принимает вид:

$$n=\frac{t^2Npq}{\Delta^2(N-1)+t^2pq}\,. \tag{27}$$

В большинстве практических случаев объем генеральной совокупности много больше единицы $N\gg 1$, что позволяет записать окончательное выражение

$$n=\frac{t^2Npq}{\Delta^2N+t^2pq}\,, \qquad N\gg 1\,. \tag{28}$$

При $\Delta^2N\gg t^2pq$ последнее выражение приобретает более простой вид

$$n=\frac{t^2pq}{\Delta^2}\,, \qquad \Delta^2N\gg t^2pq\,. \tag{29}$$

В медицинских исследдованиях удобно использовать альтернативные показатели p^*, $q^*=1000-p^*$, Δ^*, выраженные в промилле и связанные соотношением

$$p^*=1000\cdot p\,,\ q^*=1000\cdot q\,,\ \Delta^*=1000\cdot\Delta\,,\ \ p^*+q^*=1000\,. \tag{30}$$

Умножим в выражении (28) числитель и знаменатель на 10^6, получим

$$n=\frac{t^2Npq\cdot 10^6}{\Delta^2\cdot 10^6N+t^2pq\cdot 10^6}=\frac{t^2Np^*q^*}{\Delta^{*2}N+t^2p^*q^*}\,, \qquad N\gg 1\,. \tag{31}$$

В дальнейшем при расчетах будем опускать знак (*), полагая что задана соответствующая размерность.

Заключение.

Рассмотрены основные задачи, связанные с использование выборочного метода проведения исследований. Подробно исследована задача определения требуемого объема бесповторной выборки. Получена формула для определения требуемого объема выборки необходимый для получения с требуемой точностью Δ результатов при заданной вероятности γ. Показаны основные предпосылки и допущения, которые были использованы при построения окончательного результата.

Список использованных источников:

1. Гмурман В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман – Москва: Высшая школа, 1972. – 368 с.
2. Вентцель Е.С. Теория вероятностей и ее инженерные применения / Е.С. Вентцель– Москва: Высшая школа, 2000. – 480 с.

Секция ***БИЗНЕС-ИНФОРМАТИКА (Е-БИЗНЕС, ЭЛЕКТРОННЫЙ БИЗНЕС)***

РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ ПРИ ОНЛАЙН-БРОНИРОВАНИИ ГОСТИНИЧНЫХ АПАРТАМЕНТОВ

Боровых Р.И.
г. Брянск, ФГБОУ ВПО «Брянский государственный технический университет»,
09.04.04 – «Программная инженерия»

Аннотация. В данной работе исследуются алгоритмы подбора наилучшего предложения при бронировании гостиничных апартаментов через интернет-площадку. Ранжирование выдачи позволит показать пользователю наиболее релевантные предложения, как на основе истории его личных бронирований, так и с учетом анализа бронирований других пользователей.

При использовании пользователем фильтра подбора, помимо выбранных критериев, система также учитывает результаты формулы, показанной на рисунке 1.