

спіруліни в біотехнологіях харчування і фармакології: Матер. укр. наук.-практ. конф. — Вінниця, 1997. — С. 28-29.

16. *К перспективе использования спирулины в гериатрии* / Нгуен Тхиен Тхань, Л. П. Купраш, М. И. Заика и др. // *Гериатрические средства: экспериментальный поиск и клиническое использование: Тез. докл.* — К., 1990. — С. 121-122.

17. *Перспективи спіруліни в біотехнологіях харчування і фармакології: Матер. укр. наук.-практ. конф.* — Вінниця, 1997. — 89 с.

18. *Резников А. Г. Методы определения гормонов.* — К.: Наук. думка, 1980. — 400 с.

19. *Стальная И. Д., Гаришвили Т. Г. Метод определения малонового*

диальдегида с помощью тиобарбитуровой кислоты // *Современные методы в биохимии.* — М.: Медицина, 1977. — С. 66-68.

20. *Використання спіруліни в гериатрії* / І. С. Чекман, Н. О. Горчакова, Л. П. Купраш та ін. // *Матер. 3-го Нац. конгр. геронтологів і гериатрів України.* — К., 2000. — С. 36.

21. *Bhat V. B., Madyastha K. M. C-Phycocyanin: A potent peroxy radical scavenger in vivo and in vitro* // *Biochem. Biophys. Res. Commun.* — 2000. — Vol. 275, N 1. — P. 20-25.

22. *Hendrickson R. Earthfood spirulina. How this remarkable blue-green algae can transform your health and our planet.* — Laguna-Beach: Honore Enterprises, 1989. — 180 p.

23. *Howard D. The Spirulina Diet* / Lyle Stuart, Secacus N. J. — 1982. — 187 p.

24. *Matthews E. K., Saffran M. Steroid production and membrane potential measurement in cells of the adrenal gland* // *J. Physiol. (London).* — 1967. — Vol. 189, N 1. — P. 149-161.

25. *Antioxidant activity of the microalga Spirulina maxima* / M. S. Miranda, R. G. Cintra, S. B. Barros et al. // *Braz. J. Med. Biol. Res.* — 1998. — Vol. 16, N 6. — P. 1075-1079.

26. *Sautier C., Tremolieres J. Food value of spirulina in humans* // *Ann. Nutr. Alim.* — 1976. — Vol. 30. — P. 517-534.

27. *Spirulina* ETTA Nat. symp. MCRC. — Madras, India, 1992. — 130 p.

УДК 618.145-007.41:618.177

В. Н. Запорожан¹, И. З. Гладчук¹, А. П. Рогачевский¹,
В. Г. Кузнецов², Н. И. Послайко²

СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕЗУЛЬТАТОВ ГИСТЕРОСАЛЬПИНГОСЦИНТИГРАФИИ В ВЫБОРКЕ МАЛЫХ РАЗМЕРОВ С ПРИМЕНЕНИЕМ ТЕОРИИ РАСПОЗНАВАНИЯ ОБРАЗОВ И ЕВКЛИДОВОЙ МЕТРИКИ, ДОПОЛНЕННОЙ ВЕСОВЫМИ КОЭФФИЦИЕНТАМИ

¹Одесский государственный медицинский университет,

²Днепропетровский национальный университет железнодорожного транспорта
им. акад. В. Лазаряна

Гистеросальпингосцинтиграфия — новый метод обследования транспортной функции матки у женщин, страдающих бесплодием [8]. Тот факт, что он еще не получил достаточно широкого распространения, предопределяет небольшое число исследований. При проведении статистической обработки данных, полученных при обследовании бесплодных женщин с эндометриозом с применением гистеросальпингосцинтиграфии (ГССГ), перед нами стала сложная задача получения статистически достоверных результатов, учитывая, что размеры выборки были недостаточны для традиционного анализа, применяемого в медицине.

Статистический материал по исследованию женщин, больных эндометриозом, разделен на 4 группы:

1. Ипсилатеральный транспорт (7 пациенток).

2. Билатеральный транспорт (8 пациенток).

3. Контралатеральный транспорт (11 пациенток).

4. Отсутствует транспорт (17 пациенток).

Пятую группу (контрольную) составили 19 здоровых женщин из пар с только мужским фактором бесплодия.

Каждая пациентка характеризовалась 19 признаками:

1) x_{i1}^j — день цикла;

2) x_{i2}^j — размер фолликула, мм;

3) x_{i3}^j — сторона овуляции;

4) x_{i4}^j — степень эндометриоза;

5) x_{i5}^j — возраст;



- 6) x_{i6}^j — вид бесплодия;
- 7) x_{i7}^j — длительность бесплодия;
- 8) x_{i8}^j — ТТГ;
- 9) x_{i9}^j — T_3 ;
- 10) x_{i10}^j — T_4 ;
- 11) x_{i11}^j — ДГЭА-С;
- 12) x_{i12}^j — андростендион;
- 13) x_{i13}^j — 17α -ОН-прогестерон;
- 14) x_{i14}^j — ЛГ;
- 15) x_{i15}^j — ФСГ;
- 16) x_{i16}^j — ПЛ;
- 17) x_{i17}^j — E_2 ;
- 18) x_{i18}^j — тестостерон;
- 19) x_{i19}^j — ССГ.

Условные сокращения: i — номер группы

($i = \overline{1;5}$); j — номер пациентки в i -й группе;

$j = \overline{1;n_i}$ ($n_1=7, n_2=8, n_3=11, n_4=17, n_5=19$).

Исходя из наших потребностей, была сформулирована математическая постановка задачи:

1. Провести анализ признаков с точки зрения их значимости (существенности) и зависимости между собой (малосущественные признаки и дублирующие друг друга в дальнейшем можно исключить из математической модели, тем самым ее упростив).

2. Провести анализ «близости» и достоверности различий между выделенными группами пациентов.

3. Указать интервалы надежности по каждому признаку.

Описание математической модели

По имеющимся массивам данных вида

$$\begin{matrix} x_{i1}^1 & x_{i2}^1, & \dots, & x_{iN}^1 \\ x_{i1}^2 & x_{i2}^2, & \dots, & x_{iN}^2 \\ \dots, & \dots, & \dots, & \dots \\ x_{i1}^{n_i} & x_{i2}^{n_i}, & \dots, & x_{iN}^{n_i} \end{matrix} \quad (1)$$

$i = \overline{1,5}, N = 19$

были вычислены средние значения каждого признака по группам:

$$\overline{X}_{i1}, \overline{X}_{i2}, \dots, \overline{X}_{iN},$$

$$\text{где } \overline{X}_{ik} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ik}^j, n_1 = 7, n_2 = 8, n_3 = 11, n_4 = 17, n_5 = 19. \quad (2)$$

Оценки дисперсии были получены по следующим формулам:

$$S_{i1}^2, S_{i2}^2, \dots, S_{iN}^2,$$

$$\text{где } S_{ik}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ik}^j - \overline{X}_{ik})^2, \quad (3)$$

а также были определены оценки среднеквадратических отклонений:

$$S_{i1}, S_{i2}, \dots, S_{iN}, (S_{ik} = \sqrt{S_{ik}^2}).$$

Составлены матрицы эмпирических моментов корреляции:

$$K_i = [K_i(l, m)],$$

$$\text{где } K_i(l, m) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{il}^j - \overline{X}_{il})(X_{im}^j - \overline{X}_{im}). \quad (4)$$

Определены нормированные корреляционные матрицы

$$R_i = [R_i(l, m)],$$

$$\text{где } R_i(l, m) = \frac{K_i(l, m)}{S_{il} \cdot S_{im}}. \quad (5)$$

Те признаки l и m , у которых $R_i(l, m)$ по модулю близки к единице, дублируют друг друга, и из рассмотрения один из них может быть исключен. Определение значимости каждого признака по группам приведено ниже.

Решение поставленной задачи может осуществляться с помощью различных математических методов. Наиболее подходящими методами для этой цели являются методы теории распознавания образов. В дальнейшем изложении используются термины этой теории, поэтому приведем некоторые из них.

Основное понятие теории распознавания образов — образ, или класс [1]. Под образом понимают множество предметов или объектов, обладающих (объединенных) некоторыми общими свойствами. В нашем случае такими образами являются 5 групп пациентов. Каждый образ может быть представлен некоторым количеством объектов (реализаций, образцов). Описание объекта представляет собой вектор с числом компонент, равных количеству признаков, характеризующих объект [2] (в нашем случае 19 признаков), а каждая компонента задает значение соответствующего признака. Одним из наглядных методов распознавания является геометрический метод [7]. Он основан на использовании некоторой функции подобия (принадлежности) объекта S к данному классу. Эта функция определяет меру близости объекта b_j с координатами

$$\overline{x} = (y_1^m, y_2^m, \dots, y_N^m)$$

к множеству эталонов

$$\overline{y}^m = (y_1^m, y_2^m, \dots, y_n^m),$$

$$m = \overline{1, M},$$

где M — количество эталонов.



За меру близости очень часто принимается среднеквадратическое расстояние

$$S(\bar{x}, \bar{y}^m) = \frac{1}{M} \sum_{m=1}^M d^2(\bar{x}, \bar{y}^m). \quad (6)$$

Метрика d (или метод измерения расстояния) в каждом отдельном случае может быть разной. При геометрическом методе распознавания определяют такую оптимальную метрику, при которой минимизировалось бы среднеквадратическое расстояние между эталонами m -го класса [4]:

$$D^2(\bar{y}_{mi}; \bar{y}_{mj}) = \frac{1}{M_m(M_m - 1)} \times \sum_{i=1}^{M_m} \sum_{j=1}^{M_m} d^2(\bar{y}_{mi}; \bar{y}_{mj}) \rightarrow \min, \quad (7)$$

где \bar{y}_{mi} — i -й эталон m класса; M_m — количество эталонов в m -м классе.

В дальнейшем будем рассматривать такой класс метрик, который описывается формулой

$$d(\bar{a}, \bar{b}) = \sqrt{\sum_{n=1}^N \omega_{nn}^2 (a_n - b_n)^2}. \quad (8)$$

В отличие от евклидовой метрики ($d(\bar{a}, \bar{b}) = \sqrt{\sum_{n=1}^N (a_n - b_n)^2}$) придаются разные веса ω_{nn} , поскольку признаки объекта данного класса неодинаково важны при распознавании образов. Для подчеркивания существенных признаков им придается больший вес (большие значения весовых коэффициентов ω_{nn}).

При использовании метрики (7) система координат преобразуется таким образом, при котором объекты одного класса сжимаются, а множества различных классов удаляются друг от друга. При такой метрике поворот координатных осей не производится, а изменяется только масштаб вдоль координатных осей [3; 5].

Предположим, что есть один класс с M эталонами, каждый из которых характеризуется N признаками, тогда (8) можно представить в виде

$$D^2 = \frac{1}{M(M-1)} \sum_{p=1}^M \sum_{m=1}^M \sum_{n=1}^N \omega_{nn}^2 (x_{np} - x_{nm})^2 \rightarrow \min. \quad (9)$$

Для нахождения минимума D^2 накладывают дополнительные условия на весовые коэффициенты, например,

$$\sum_{n=1}^N \omega_{nn} = 1, \quad (10)$$

а (9) – (10) представляет собой задачу нахождения условного экстремума. С помощью функции Лагранжа ее сводят к задаче отыскания абсолютного экстремума [6]. Решение (9) – (10) имеет вид:

$$\omega_{nn} = \frac{1}{S_n^2 \sum_{n=1}^N \frac{1}{S_n^2}}, \quad (11)$$

где S_n^2 — оценка дисперсии n -го признака.

Если дисперсия в значениях каждого признака эталонов каждого класса велика, то этот признак можно исключить из рассмотрения, так как он ничего не дает при распознавании и является несущественным.

Показано, что если, кроме сжатия по осям, допустить вращение, то оптимальное положение новых осей (в смысле минимума среднего квадрата расстояния) выбирается в зависимости собственных значений корреляционной матрицы данного класса.

Решающее правило для геометрического распознавания формулируется следующим образом.

Пусть даны два класса A_1 и A_2 . Для распознавания принадлежности какого-нибудь объекта \bar{p} к классу $\{x_m\}$ определяется величина

$$S(\bar{p}, \{x_m\}) = \frac{1}{M} \sum_{m=1}^M \sum_{n=1}^N \left[\sum_{s=1}^N a_{ns} (p_s - x_{ms}) \right]^2. \quad (12)$$

Если $S_{A_1}(\bar{p}, \{x_m\}) < S_{A_2}(\bar{p}, \{x_m\})$, то $p \in A_1$,

если же $S_{A_1}(\bar{p}, \{x_m\}) > S_{A_2}(\bar{p}, \{x_m\})$, то $p \in A_2$,

причем в каждой из функций S_{A_1} и S_{A_2} коэффициенты a_{ns} ищутся применительно к первому или второму множеству эталонов.

После анализа весовых коэффициентов по всем классам можно выбрать какие-то общие для всех коэффициенты преобразования с учетом, например, априорных вероятностей появления классов.

Запишем вычислительные формулы для весовых коэффициентов и решающее правило в обозначениях нашей задачи.

Пусть $S(\bar{p}, A_i)$ — среднеквадратическое расстояние от вектора $\bar{p} = (p_1, p_2, \dots, p_N)$, который представляет объект, подлежащий распознаванию, к элементам образа A_i , тогда

$$S(\bar{p}, A_i) = \frac{1}{n_i} \sum_{m=1}^{n_i} \sum_{n=1}^N \omega_{ni}^2 (p_n - x_{in}^m)^2, \quad i = \overline{1,5}. \quad (13)$$

Решающее правило в нашем случае будет выглядеть так: пациент \bar{p} относится к группе пациентов A_k , если, $S(\bar{p}, A_k) = \min S(\bar{p}, A_i), 1 \leq i \leq 5$. (14)



Средние значения, дисперсии, весовые коэффициенты признаков

Номер признака n	\bar{X}_{1n}	\bar{X}_{2n}	\bar{X}_{3n}	\bar{X}_{4n}	\bar{X}_{5n}	S_{1n}^2	S_{2n}^2	S_{3n}^2	S_{4n}^2	S_{5n}^2	ω_{n1}	ω_{n2}	ω_{n3}	ω_{n4}	ω_{n5}
1	12,23	13,63	13,09	12,94	12,89	2,78	1,48	2,99	5,70	4,62	0,006	0,11	0,0047	0,01	0,0028
2	15,07	17,31	17,65	15,16	17,37	14,07	19,56	12,43	8,55	8,02	0,00119	0,00082	0,0011	0,055	0,0016
3	1,57	1,13	1,68	1,59	1,53	0,24	0,36	0,19	0,226	0,25	0,07	0,0446	0,0736	0,24	0,051
4	2	1,63	2,8	1,82	-	1,14	1,23	0,876	0,85	-	0,0146	0,0131	0,016	0,064	-
5	31,57	30,38	33,45	33,94	33,58	7,96	13,23	10,25	11,70	16,56	0,0021	0,0012	0,00136	0,0047	0,0078
6	1,42	1	1,09	1,29	1,26	0,24	0	0,083	0,21	0,19	0,07	-	0,169	0,26	0,0676
7	31,57	4,63	4,14	4,62	3,58	7,96	5,48	3,69	6,81	2,64	0,0021	0,0029	0,0038	0,008	0,0049
8	1,23	1,88	2,11	1,484	2,83	0,53	0,80	0,69	0,85	4,05	0,0315	0,02	0,0203	0,064	0,0032
9	3,40	3,68	3,32	3,42	3,395	0,09	0,03	0,187	0,365	0,07	0,186	0,5352	0,0748	0,149	0,1784
10	1,27	1,36	1,19	2,02	1,205	0,03	0,05	0,023	1,084	0,02	0,557	0,3211	0,608	0,05	0,6422
11	1495,14	1948,8	1417,09	1660,88	1613,68	428824,69	436498,96	164558,81	475131,63	738910	$3,89 \cdot 10^{-8}$	$3,68 \cdot 10^{-8}$	$8,5 \cdot 10^{-8}$	$1,15 \cdot 10^{-7}$	$1,35 \cdot 10^{-6}$
12	3,36	3,2	1,93	2,38	2,22	0,36	0,48	0,575	0,457	0,348	0,046	0,033	0,0243	0,119	0,0369
13	596,43	603,58	595,45	604,18	421,26	39883,96	55497,71	67175,16	90182,62	23664,4	$4,18 \cdot 10^{-7}$	$2,89 \cdot 10^{-7}$	$2,08 \cdot 10^{-7}$	$6,1 \cdot 10^{-7}$	$5,43 \cdot 10^{-8}$
14	5,11	6,29	7,06	10,26	5,56	5,87	8,03	5	295,04	3,30	0,0028	0,002	0,0028	0,0188	0,0039
15	5,41	4,98	9,75	7,95	7,17	2,78	1,13	56,9	15,70	2,08	0,006	0,0142	0,000246	0,0035	0,0062
16	10,11	180,40	16,39	45,71	67,49	3,15	77476,17	49,46	6018,95	10975,1	0,0053	$2,07 \cdot 10^{-7}$	0,00028	$9,1 \cdot 10^{-6}$	$1,17 \cdot 10^{-6}$
17	63,85	83,29	33,8	88,37	62,95	1063,26	2212	298,4	5269,2	1296,79	$1,57 \cdot 10^{-5}$	$7,26 \cdot 10^{-6}$	$4 \cdot 10^{-5}$	$1,04 \cdot 10^{-5}$	$9,9 \cdot 10^{-6}$
18	41,86	46,50	40,09	34,93	31,74	40,41	50	250,63	182,67	91,14	0,0004	0,00032	$5,60 \cdot 10^{-5}$	0,0003	0,00014
19	74,71	63	67,55	68,94	53,74	482,2	255,08	209,34	1241,35	214,4	$3,46 \cdot 10^{-5}$	$6,3 \cdot 10^{-5}$	$6,7 \cdot 10^{-5}$	$4,4 \cdot 10^{-5}$	$6 \cdot 10^{-5}$



Таблица 2

Корреляционная матрица

n \ n	3	4	6	8	9	10	12
3	1	0,27	-0,33	-0,04	-0,19	-0,47	-0,58
4		1	-0,27	-0,35	0,66	0,23	-0,21
6			1	0,05	-0,38	-0,52	0,12
8				1	-0,72	0,11	-0,1
9					1	0,54	-0,42
10						1	-0,84
12							1

Статистический материал по пяти группам пациентов для удобства представлен в виде таблиц. Признак с номером 3 (сторона овуляции) не числовой, поэтому его значениям поставлены в соответствие числа: «слева» — 2, «справа» — 1, «с обеих сторон» — 0. В табл. 1 приведены средние значения, дисперсии, весовые коэффициенты признаков по всем пяти группам пациентов. Для пятой группы пациентов 4-й признак (степень эндометриоза) отсутствует, поэтому в соответствующих ячейках таблицы стоят прочерки. Кроме того, во 2-й группе пациентов признак с номером 6 (вид бесплодия) оказался постоянным, его дисперсия равна нулю, поэтому его весовой коэффициент не определен, что тоже отмечено прочерком.

По данным табл. 1 был проведен анализ признаков с точки зрения их значимости (существенности) при распознавании заболевания. Оказалось, что из 19 признаков наиболее существенными являются семь с номерами 3, 4, 6, 8, 9, 10, 12. Сумма их весовых коэффициентов по всем классам составляет в среднем 97 % от общей суммы.

Среди выделенных семи наиболее существенных признаков практически нет дублирующих. Это видно по матрицам коэффициентов корреляции по всем пяти группам пациентов. В табл. 2 приведена корреляционная матрица по первой группе пациентов. Согласно данным этой таблицы, наиболее коррелированными являются 10-й и 12-й признаки, представляющие собой уровень тетраодтиронина и андростендиона в сыворотке крови пациенток. Можно предположить, что Т4 — гормон, реализующий периферические метаболические эффекты и влияющий на рецепторогенез, а андростендион — продукт метаболизма половых стероидов (в основном тестостерона), поэтому изменение его уровня отражает степень нарушения гормонального статуса, модулирующего сократительную способность субэндометриального слоя миометрия. Приведенная интерпретация полученных результатов — гипотетическая, требующая дальнейших исследований.

Таким образом, созданная нами методика статистической обработки данных при малых размерах выборок может использоваться и в медицине, особенно в тех ситуациях, когда невозможно собрать достаточное для традиционных методов статистики количество пациентов.

ЛИТЕРАТУРА

1. Вапник В. Н., Червоненкис А. Я. Теория распознавания образов. Стохастические проблемы обучения. — М.: Наука, 1974.
2. Боровков А. А. Математическая статистика. — М.: Наука, 1984.
3. Андерсон Т. Введение в многомерный статистический анализ: Пер. с англ. — М.: Физматгиз, 1963.
4. Вапник В. Н. Восстановление зависимостей по эмпирическим данным. — М.: Наука, 1979.
5. Патрик Э. Основы теории распознавания образов: Пер. с англ. / Под ред. Б. Р. Левина. — М.: Сов. радио, 1980.
6. Справочник по теории вероятностей и математической статистике / В. С. Королюк, Н. И. Портенко, А. В. Скороход, А. Ф. Турбин. — М.: Наука, 1985.
7. Ту Дж., Гонсалес Р. Принципы распознавания образов: Пер. с англ. — М.: Мир, 1978.
8. Рогачевський О. П. Чи існує взаємозв'язок між транспортною функцією матки та фолікулогенезом у безплідних жінок з ендометріозом? // Шпит. хірургія. — 2005. — № 2. — С. 47-51.

