* 1. Разработка дискриминантных функций для диагностики легочных заболеваний

Для поиска переменных, позволяющих относить наблюдаемые объекты в группы использовался дискриминантный анализ. Главной целью дискриминации является нахождение такой линейной комбинации переменных, которая бы оптимально разделила бы рассматриваемые группы.

На первом этапе проведения дискриминантного анализа разделили исходную выборку на две части - обучающую и тестовую. В тестовую вошли 24 пациента (по 5 с диагнозом кавернозный, инфильтративный и диссеминированный туберкулез, 4 пациента с очаговым туберкулезом, 3 с пневмонией и 2 с онкологией). Остальные пациенты составили обучающую выборку (73 пациента), которую и использовали для вычисления дискриминантных функций.

В процессе проведения дискриминантного анализа объединили признаки в группы. Были рассмотрены следующие группы:

- «Анамнез», в которую включены такие признаки как слабость, утомляемость, потеря веса, бледность, температура, кашель, мокрота, кровохаркание, рвота, головные боли, кровь в каловых массах, поносы и боли в груди;

- «Вид мокроты» (гнойная, слизистая или кровяная);

- «Анализ крови» (включает такие показатели как повышенное содержание эритроцитов, лейкоцитов, лимфоцитов, гемоглобин и СОЭ);

- «Размер тени» (мелкие, средние, крупные);

- «Локализация затемнения» (верхние и нижние отделы, одинарные и множественные затемнения);

- «Стенки полости» (гладкие, ровные, четкие, нечеткие, бугристые, рыхлые, отечные, неоднородные);

- «Форма» (щелевидная, округлая, кольцевидная, овальная, неправильная, облаковидная, звездчатая);

- «Наличие уплотнений».

Далее выбрали в программе STATISTICA такую процедуру многомерного разведочного анализа как дискриминантный анализ. В качестве группируемой переменной выбрали «Диагноз», а остальные признаки и показатели исследований - независимые переменные. После определения анализируемой выборки вычисляются коэффициенты дискриминантной функции.

Итоги анализа дискриминантных функций представлены в таблице 3. По полученной таблице можем сделать вывод о том, что по р – уровню значимости все критерии значимы (р –уровень < 0,05).

Таблица 3 – Итоги анализа дискриминантных функций (Переменных в модели: 8; Групп.: диагноз (6 гр.) Лямбда Уилкса: 0,03408 прибл. F (40,259)=7,6107 p<0,0000)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Уилкса | Частная | F-ключ | p-уров. | Толер. | 1-толер. |
| Анамнез | 0,056283 | 0,605537 | 8,68684 | 0,000013 | 0,878062 | 0,121939 |
| Вид мокроты | 0,035874 | 0,850028 | 2,02069 | 0,049451 | 0,882606 | 0,117394 |
| Анализ крови | 0,074094 | 0,459972 | 13,85372 | 0,000000 | 0,981423 | 0,018577 |
| Размер тени | 0,047468 | 0,717980 | 5,63500 | 0,001236 | 0,921997 | 0,078003 |
| Локализация затемнения | 0,054473 | 0,625651 | 7,86036 | 0,000031 | 0,916051 | 0,083949 |
| Стенки полости | 0,039985 | 0,752344 | 3,04417 | 0,035454 | 0,987568 | 0,012432 |
| Форма | 0,065070 | 0,523766 | 10,72915 | 0,000000 | 0,962855 | 0,037145 |
| Наличие уплотнений | 0,037747 | 0,802879 | 2,26931 | 0,039251 | 0,929369 | 0,070631 |

Λ - статистика Уилкса (лямбда статистика) является стандартной статистикой, используемой для обозначения статистической значимости мощности дискриминации в текущей модели. Ее значение меняется от 1.0 (нет никакой дискриминации) до 0.0 (полная дискриминация). Чем меньше Λ - статистика Уилкса, тем значимее дискриминантная функция. Таким образом видим, что наиболее значимой является переменная «Вид мокроты», так как Λ - статистика Уилкса в данном случае равна 0,035874, что является наименьшим значением среди остальных.

Частная Λ - статистика Уилкса – это статистика для одиночного вклада соответствующей переменной в дискриминацию между совокупностями. Так как лямбда с величиной 0.0 обозначает полную дискриминацию, то чем ниже ее значение в этом столбце, тем больше одиночный вклад соответствующей переменной в степень дискриминации. Таким образом наибольший одиночный вклад вносит переменная «Анализ крови» (частная Λ – Уилкса равна 0,459972) и «Форма» (частная Λ – Уилкса равна 0,523766).

F - исключ. и p - значение считается также как и F - статистика для всей модели, только вместо Λ Уилкса подставляется частная Λ Уилкса. Применительно к нашей таблице наибольшие значения величины F- исключ. дает переменная «Анализ крови» (F - исключ. = 13,85372) и «Форма» (F - исключ. = 10,72915), т.е. она вносит наибольший вклад в модель.

Значение толерантности определяется как 1 минус R-квадрат для соответствующей переменной со всеми другими переменными в модели. Это значение дает информацию об избыточности данной переменной. Когда другие переменные еще не выбраны, все R-квадрат равны 1.0.

В дискриминантном анализе есть возможности определить дискриминантные функции таким образом, чтобы в эти функции вошли только статистически значимые переменные. Оценку надежности и качества дискриминантных функций позволяет произвести классификационная матрица (таблица 4).

Таблица 4 – Классификационная матрица

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Процент | Кавернозный | Инфильтративный | Очаговый | Диссеминированный | Онкология | Пневмония |
| Кавернозный | 57,1429 | 8 | 1 | 2 | 3 | 0 | 0 |
| Инфильтративный | 80,0000 | 1 | 12 | 2 | 0 | 0 | 0 |
| Очаговый | 72,7273 | 1 | 1 | 8 | 1 | 0 | 0 |
| Диссеминированный | 86,6667 | 0 | 1 | 0 | 13 | 0 | 1 |
| Онкология | 85,7143 | 0 | 0 | 1 | 0 | 6 | 0 |
| Пневмония | 100,0000 | 0 | 0 | 0 | 0 | 0 | 10 |
| Всего | 79,1667 | 10 | 15 | 13 | 17 | 6 | 11 |

В полученной таблице в строках представлены наблюдаемые группы, а в столбцах – предсказанные. По данной таблице мы можем посмотреть процент корректной классификации для каждой группы.

Так для первой группы процент корректной классификации пациентов с диагнозом кавернозный туберкулез составил 57,1%, для второй (инфильтративный туберкулез) – 80%, для третьей (очаговый туберкулез) – 72,7%, для четвертой (диссеминированный туберкулез) – 86,7%, для пятой (онкология) – 79,2 %, для шестой (пневмония) – 100%. В среднем процент корректной классификации по всем группам составил 79,2%.

Расстояние Махаланобиса является мерой расстояния, которую можно использовать в многомерном пространстве, определенном переменными модели и позволяет вычислить расстояние между наблюдением и центром каждой совокупности. Также можно непосредственно вычислить вероятность того, что наблюдение принадлежит определенной совокупности. Это условная вероятность в том смысле, что она зависит от знания значений переменных в модели. Поэтому эти вероятности называются апостериорными вероятностями.

В таблице 5 представлены квадраты расстояний Махаланобиса. Символом (\*) помечены некорректно отнесенные пациенты (в нашем случае были помечены: 102, 107, 112,113,119,120,202,204,205, 305,312,313, 401,402,404,508).

Таблица 5 - Квадраты расстояний Махаланобиса до центров

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Наблюд. | Кавернозный | Инфильтративный | Очаговый | Диссеминированный | Онкология | Пневмония |
| \*пациент 102 | Кавернозный | 10,04274 | 14,06369 | 8,06070 | 23,05794 | 10,50849 | 38,02663 |
| пациент 103 | Кавернозный | 4,11221 | 15,94050 | 6,75447 | 10,09248 | 13,57111 | 25,01157 |
| пациент 105 | Кавернозный | 4,16905 | 18,26870 | 11,25313 | 12,34344 | 21,13345 | 29,16858 |
| \*пациент 107 | Кавернозный | 8,59419 | 6,15647 | 12,45760 | 15,83874 | 22,76850 | 21,68540 |
| пациент 108 | Кавернозный | 14,03046 | 15,67873 | 22,05233 | 21,47615 | 40,34086 | 28,00996 |
| пациент 109 | Кавернозный | 6,40019 | 14,71925 | 9,44001 | 7,29392 | 25,50035 | 10,15748 |
| пациент 110 | Кавернозный | 5,40366 | 11,89980 | 6,31896 | 17,29579 | 9,50222 | 32,22423 |
| пациент 111 | Кавернозный | 2,93041 | 11,34657 | 5,73778 | 5,73082 | 15,67962 | 25,07780 |
| \*пациент 112 | Кавернозный | 5,76704 | 14,80093 | 13,13911 | 4,75204 | 15,26356 | 19,36646 |
| \*пациент 113 | Кавернозный | 6,71091 | 8,51904 | 10,90637 | 6,62192 | 23,06489 | 13,64966 |
| пациент 116 | Кавернозный | 3,81985 | 14,21744 | 7,55424 | 14,36108 | 18,77055 | 18,03575 |
| пациент 118 | Кавернозный | 4,08950 | 16,06145 | 8,95449 | 14,42616 | 19,58384 | 30,53335 |
| \*пациент 119 | Кавернозный | 10,80058 | 14,71194 | 16,09561 | 7,22825 | 30,91838 | 13,48359 |
| \*пациент 120 | Кавернозный | 7,36358 | 6,01117 | 2,85240 | 11,88526 | 15,38759 | 30,07358 |
| пациент 201 | Инфильтративный | 13,85017 | 7,66243 | 15,28977 | 22,63271 | 20,47306 | 32,04063 |
| \*пациент 202 | Инфильтративный | 5,39963 | 6,52035 | 4,38475 | 6,64481 | 15,75588 | 17,84304 |
| пациент 203 | Инфильтративный | 5,99288 | 4,58683 | 5,53019 | 8,80886 | 8,26933 | 26,10946 |
| \*пациент 204 | Инфильтративный | 9,75469 | 11,20882 | 10,87603 | 14,30466 | 32,30379 | 22,09483 |
| \*пациент 205 | Инфильтративный | 5,80635 | 5,70014 | 4,35161 | 5,94115 | 17,88489 | 19,41691 |
| пациент 206 | Инфильтративный | 16,15436 | 4,93609 | 10,23137 | 23,64488 | 17,72875 | 50,99402 |
| пациент 207 | Инфильтративный | 12,27089 | 3,24283 | 10,64486 | 21,86492 | 16,96358 | 30,22510 |
| пациент 208 | Инфильтративный | 16,78361 | 12,81755 | 13,54031 | 25,76713 | 37,26475 | 36,13642 |
| пациент 212 | Инфильтративный | 13,26227 | 3,67137 | 11,87232 | 22,32888 | 19,03860 | 29,92134 |
| пациент 214 | Инфильтративный | 16,54784 | 3,70745 | 13,66722 | 22,63506 | 23,02188 | 43,36937 |
| пациент 215 | Инфильтративный | 35,87554 | 21,75212 | 24,20423 | 35,62587 | 33,72929 | 62,16585 |
| пациент 216 | Инфильтративный | 11,21022 | 4,78244 | 9,67878 | 9,41668 | 18,59609 | 36,85183 |
| пациент 218 | Инфильтративный | 9,80750 | 5,90142 | 9,32570 | 20,81979 | 15,03634 | 31,57007 |
| пациент 219 | Инфильтративный | 12,99090 | 5,23972 | 12,33601 | 9,92669 | 21,92510 | 36,14667 |
| пациент 220 | Инфильтративный | 15,49156 | 5,01029 | 8,81261 | 19,77349 | 27,95169 | 42,53006 |
| пациент 303 | Очаговый | 11,83705 | 23,94911 | 10,33815 | 11,46903 | 26,31176 | 29,08873 |
| \*пациент 305 | Очаговый | 7,13317 | 5,22248 | 5,18791 | 17,50803 | 14,44849 | 20,33548 |
| пациент 307 | Очаговый | 9,73744 | 11,43845 | 6,02782 | 10,87993 | 23,91045 | 31,51339 |
| пациент 308 | Очаговый | 13,99067 | 11,52757 | 8,02290 | 19,92286 | 9,33847 | 32,11668 |
| пациент 309 | Очаговый | 10,26538 | 7,35275 | 6,64076 | 16,32913 | 17,49956 | 24,52663 |
| пациент 310 | Очаговый | 8,91949 | 6,22226 | 5,47323 | 18,04876 | 5,19944 | 33,91744 |
| пациент 311 | Очаговый | 9,48794 | 5,59112 | 3,84753 | 12,89849 | 21,38642 | 31,70552 |
| \*пациент 312 | Очаговый | 11,80144 | 12,74493 | 10,14195 | 7,53735 | 28,30259 | 34,13300 |
| \*пациент 313 | Очаговый | 4,42937 | 11,69419 | 8,80979 | 9,90102 | 20,82798 | 11,43760 |
| пациент 314 | Очаговый | 15,63216 | 11,26310 | 5,54242 | 19,92862 | 16,70197 | 45,06717 |
| пациент 315 | Очаговый | 8,36245 | 9,93032 | 4,36617 | 20,72301 | 11,85968 | 34,35888 |
| \*пациент 401 | Диссеминированный | 10,46808 | 6,71582 | 6,35417 | 12,64403 | 22,99656 | 41,03714 |
| пациент 402 | Диссеминированный | 12,15363 | 19,11036 | 16,27757 | 5,90135 | 30,82629 | 35,22965 |
| \*пациент 404 | Диссеминированный | 19,26778 | 26,48873 | 18,56645 | 18,60618 | 27,02440 | 8,46685 |
| пациент 407 | Диссеминированный | 29,07236 | 33,92546 | 30,07233 | 14,06480 | 31,05102 | 26,25197 |
| пациент 408 | Диссеминированный | 10,45168 | 13,46598 | 17,01582 | 5,87435 | 29,81433 | 33,67475 |
| пациент 409 | Диссеминированный | 5,59428 | 12,85588 | 12,52923 | 5,47490 | 28,32964 | 25,54455 |
| пациент 410 | Диссеминированный | 14,39164 | 17,03148 | 19,73433 | 8,49604 | 29,84877 | 21,17424 |
| пациент 411 | Диссеминированный | 23,36564 | 22,37573 | 18,54239 | 15,22527 | 23,35308 | 29,38598 |
| пациент 412 | Диссеминированный | 7,30298 | 21,83530 | 14,99744 | 5,92409 | 20,82843 | 23,07237 |
| пациент 413 | Диссеминированный | 13,05049 | 24,74701 | 19,38187 | 5,44672 | 29,52884 | 21,96370 |
| пациент 415 | Диссеминированный | 23,87745 | 35,48947 | 30,65515 | 11,01794 | 47,21435 | 40,59356 |
| пациент 416 | Диссеминированный | 7,07678 | 11,36107 | 11,50100 | 3,84235 | 28,99241 | 30,42524 |
| пациент 417 | Диссеминированный | 33,10261 | 32,59436 | 26,96933 | 20,07492 | 40,06271 | 30,45305 |
| пациент 418 | Диссеминированный | 12,04205 | 18,85182 | 17,59226 | 11,83724 | 36,21317 | 19,48904 |
| пациент 419 | Диссеминированный | 12,05361 | 16,06519 | 13,07532 | 5,08776 | 34,98626 | 32,76556 |
| пациент 501 | Онкология | 15,81617 | 11,87053 | 8,80044 | 22,97250 | 2,70138 | 38,59109 |
| пациент 502 | Онкология | 16,99426 | 23,24035 | 14,52614 | 28,09076 | 8,37097 | 35,65457 |
| пациент 503 | Онкология | 44,14866 | 44,71463 | 47,13792 | 49,39276 | 17,36649 | 57,20792 |
| пациент 506 | Онкология | 19,84446 | 15,27712 | 17,00977 | 20,72823 | 6,10331 | 41,11827 |
| \*пациент 508 | Онкология | 6,20452 | 9,02424 | 5,22295 | 14,90430 | 8,14306 | 16,99734 |
| пациент 509 | Онкология | 24,05585 | 27,85951 | 22,72916 | 30,43288 | 3,49910 | 39,58903 |
| пациент 510 | Онкология | 20,69161 | 24,18446 | 15,69669 | 32,74569 | 7,69172 | 33,04744 |
| пациент 601 | Пневмония | 24,68134 | 26,14017 | 23,92049 | 15,77871 | 41,88293 | 9,73844 |
| пациент 602 | Пневмония | 20,23476 | 20,39848 | 22,23935 | 21,00842 | 31,23570 | 4,75176 |
| пациент 603 | Пневмония | 28,10434 | 42,60236 | 35,25382 | 31,81822 | 31,80575 | 4,07250 |
| пациент 604 | Пневмония | 19,17683 | 32,66157 | 25,85144 | 18,73780 | 37,62072 | 3,88646 |
| пациент 606 | Пневмония | 12,73774 | 21,69195 | 19,00679 | 11,72870 | 36,87078 | 5,06782 |
| пациент 607 | Пневмония | 18,87180 | 30,04791 | 24,89808 | 18,89272 | 26,48950 | 2,68085 |
| пациент 608 | Пневмония | 15,16681 | 26,74351 | 22,17824 | 18,48673 | 30,57187 | 2,35646 |
| пациент 609 | Пневмония | 33,08226 | 47,78866 | 41,06138 | 36,32704 | 45,05528 | 6,47377 |
| пациент 612 | Пневмония | 36,54790 | 48,21540 | 43,54211 | 39,62948 | 41,09995 | 6,23967 |
| пациент 613 | Пневмония | 12,25058 | 26,72700 | 22,01047 | 16,83181 | 24,21444 | 3,96544 |

После построения матрицы классификации можем посмотреть таблицу классификационных функций и записать их (таблица 6).

Функции классификации предназначены для отнесения объекта к наиболее вероятной группе, и их число соответствует количеству групп.

Таблица 6 – Классификационные функции

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Кавернозный | Инфильтративный | Очаговый | Диссеминированный | Онкология | Пневмония |
| Анамнез | 0,8317 | 0,6910 | 0,8907 | 0,6998 | 1,1026 | 0,5079 |
| Вид  мокроты | -0,1794 | -0,4665 | -0,3883 | -0,1103 | -0,6665 | -0,0422 |
| Анализ  крови | 1,6580 | 0,9706 | 1,3705 | 2,3850 | 1,2781 | 2,2101 |
| Размер  тени | 3,2724 | 4,0095 | 4,1512 | 3,9948 | 5,3228 | 4,4072 |
| Локализация  затемнения | 6,6819 | 7,2931 | 7,1243 | 7,3588 | 7,8928 | 5,1064 |
| Стенки  полости | 0,7874 | 0,8862 | 0,5852 | 0,8953 | 1,0969 | 0,7352 |
| Форма | 0,5056 | 0,0328 | 0,2266 | 0,0823 | 1,0958 | 1,4130 |
| Наличие  уплотнений | 0,2059 | 0,2068 | -0,1462 | -0,1409 | 0,2534 | 0,0070 |
| Константа | -58,1604 | -57,7186 | -61,4562 | -68,7652 | -89,5891 | -54,3050 |

Таким образом дискриминантные функции имеют следующий вид:

Y1 = -58.1604 + 0.8317\*X1 -0.1794\*X2 + 1.6580\*X3 +

+ 3.2724\*X4 + 6.6819\*X5 + 0.7874\*X6 + 0.5056\*X7 + 0.2059 \*X8.

Y2 = -57.7186+ 0.6910\*X1 -0.4665\*X2 + 0.9706\*X3 +

+ 4.0095\*X4 + 7.2931\*X5 + 0.8862\*X6+ + 0.0328\*X7 + 0.2068\*X8.

Y3 = -61.4562 + 0.8907\*X1 -0.3883\*X2 + 1.3705\*X3 +

+ 4.1512\*X4 + 7.1243\*X5 + 0.7874\*X6 + 0.2266\*X7 - 0.1462\*X8.

Y4 = -68.7652+ 0.6998\*X1 - 0.1103\*X2 + 2.3850\*X3 +

+ 3.9948\*X4 + 7.3588\*X5 + 0.8953\*X6 + 0.0823\*X7-0.1409\*X8.

Y5 = -89.5891 + 1.1026\*X1 -0.6665\*X2 + 1.2781\*X3 +

+ 5.3228\*X4 + 7.8928\*X5 + 1.0969\*X6 + 1.0958\*X7 + 0.2534\*X8.

Y6 = -54.3050+ 0.5079\*X1 -0.0422\*X2 + 2.2101\*X3 +

+ 4.4072\*X4 + 5.1064\*X5 + 0.7352\*X6 + 1.4130\*X7 + 0.0070\*X8.

где - результат показателя классификации,

- константа для совокупности,

– веса для j-ой переменной при вычислении показателя классификации для i-ой совокупности,

– наблюдаемое значение для соответствующего образца j-ой переменной,

индексы 1, 2, …, m – число переменных.