

УДК 159.98(075)

ДИСКРИМИНАЦИЯ, КАК СРЕДСТВО МОДЕЛИРОВАНИЯ ТРУДООХРАННЫХ МЕТРОПРИЯТИЙ

Жижин К.С.

*ГБОУ СПО РО «Ростовский базовый медицинский колледж»,
Ростов-на-Дону, e-mail: ZIZIN2007@mail.ru*

В работе рассматриваются приемы дискриминации признаков производственных травм с использованием модуля «Дискриминантный анализ» статистического софта «Statistica» v.6. Отражена простота анализа и получения выводов. Рекомендации могут быть реализованы специалистами, чей математический багаж не превышает базиса средней общеобразовательной школы.

Ключевые слова: трудовой процесс, дискриминация травмы

DISCRIMINATION, AS MEANS MODELLING OF SECURITY ACTIONS

Zhizhin K.S.

The Rostov base medical college, Rostov-on-Don, e-mail: ZIZIN2007@mail.ru

In work receptions of discrimination of attributes of industrial traumas with use of the module «the Discriminantal analysis» statistical software «Statistica» v.6 are considered. Simplicity of the analysis and reception of conclusions is reflected. Recommendations can be realized by experts, whose mathematical luggage does not exceed basis of an average comprehensive school.

Keywords: labor process, discrimination of a trauma

Трудоохранные мероприятия в современных условиях не могут не базироваться на переработке достаточно плотного потока статистических данных. Без них невозможно прогнозировать ситуацию, стоять стратегии улучшения условий труда. В череде этих проблем самая существенная – классифицирование производственных травм, профессиональных заболеваний и отравлений при формировании динамического ряда. Лет 15–20 назад такая работа требовала привлечения труда профессионалов-программистов, занимала много времени на обработку и анализ полученных результатов. Современные статистические софты в значительной мере облегчили эту работу специалистам трудоохранных служб, поскольку подоплека их интуитивно понятна даже человеку с математическим базисом на уровне 10–11 классов средней общеобразовательной школы.

Материалы и методы исследования

Дискриминантный анализ, применение которого мы демонстрируем в данном сообщении – достаточно сложный раздел математической статистики. И, тем не менее, с помощью модуля «Дискриминантный анализ» из американского статистического софта «Statistica» v.6 мы хотели бы показать насколько просто провести процесс дискриминации. Примером в данном сообщении служит классификация производственных травм по тяжести.

Результаты исследований и их обсуждение

Данная выборка включает 11 единиц наблюдений, отобранных случайным образом

из совокупности в 100 единиц. Травмы будем классифицировать, опираясь на следующие дискриминационные признаки: количество дней нетрудоспособности работника из-за одной травмы, число травм, случившихся у него в течение года, расходы на лечение в тыс. руб. (в расчете на одну травму), индекс травмирования, то есть отношение числа травм к числу дней нетрудоспособности, табл. 1. Подразумевается, что программа «Statistica» v.6. уже установлена, поэтому обходим процедуру её инсталляции на жесткий диск компьютера.

На верхней панели окна щелкаем левой кнопкой «мыши» на слове **Анализ**, отыскиваем **Многомерный разведочный анализ**, в нем – **Дискриминантный анализ** (рис. 1).

Во вкладке **Быстрый** выбираем **Дополнительные параметры** (пошаговый анализ). После нажатия на кнопку **Переменные** отобразится стандартное диалоговое окно **Выбор переменных** (его мы не показываем) (рис. 2). В этом окне укажем группирующую переменную и независимые переменные, которые должны быть использованы для дискриминации типа травм. В нашем случае *группирующим признаком будет тяжесть травмы*.

Нажимаем кнопку **ОК**, и переходим к следующему этапу: **Результаты...**, рис. 3. Просмотр результатов дискриминантного анализа, и классификация наблюдений начинаются с верхней части. В белом прямоугольнике, представлены значения самого существенного показателя дискриминации –

лямбды Уилкса, пределы её изменений: 0–1. В нашем случае значение лямбды достаточно мало – 0,0026 (Суть в том, что, если это значение близко к нулю, то дискриминация прошла успешно, если же близко к единице, то дискриминация со-

мнительна) (рис. 3). Помимо этого, полученный в опыте, своеобразный показатель достоверности вывода, критерий Фишера «F» также высок – 23,2, почти в три раза перекрывает свое стандартное значение – 8,10 (в скобках).

Таблица 1

Классификация травм

№ п/п	Var1 Дни	Var2 Случаи	Var3 Стоимость лечения	Var4 случаи/дни	Var5 Тяжесть травмы
1	50	3	1,2	0,06	Легкая
2	50	3	1,4	0,06	Легкая
3	64	2	5,6	0,05	Средней тяжести
4	65	2	4,8	0,05	Тяжелая
5	67	3	5,6	0,04	Средней тяжести
6	63	3	5,7	0,05	Средней тяжести
7	46	4	1,4	0,06	Легкая
8	69	3	5,1	0,04	Средней тяжести
9	62	2	4,5	0,03	Тяжелая
10	59	3	4,8	0,05	Тяжелая
11	45	4	1,3	0,08	Легкая

Многомерный разведочный анализ

Дискриминантный анализ

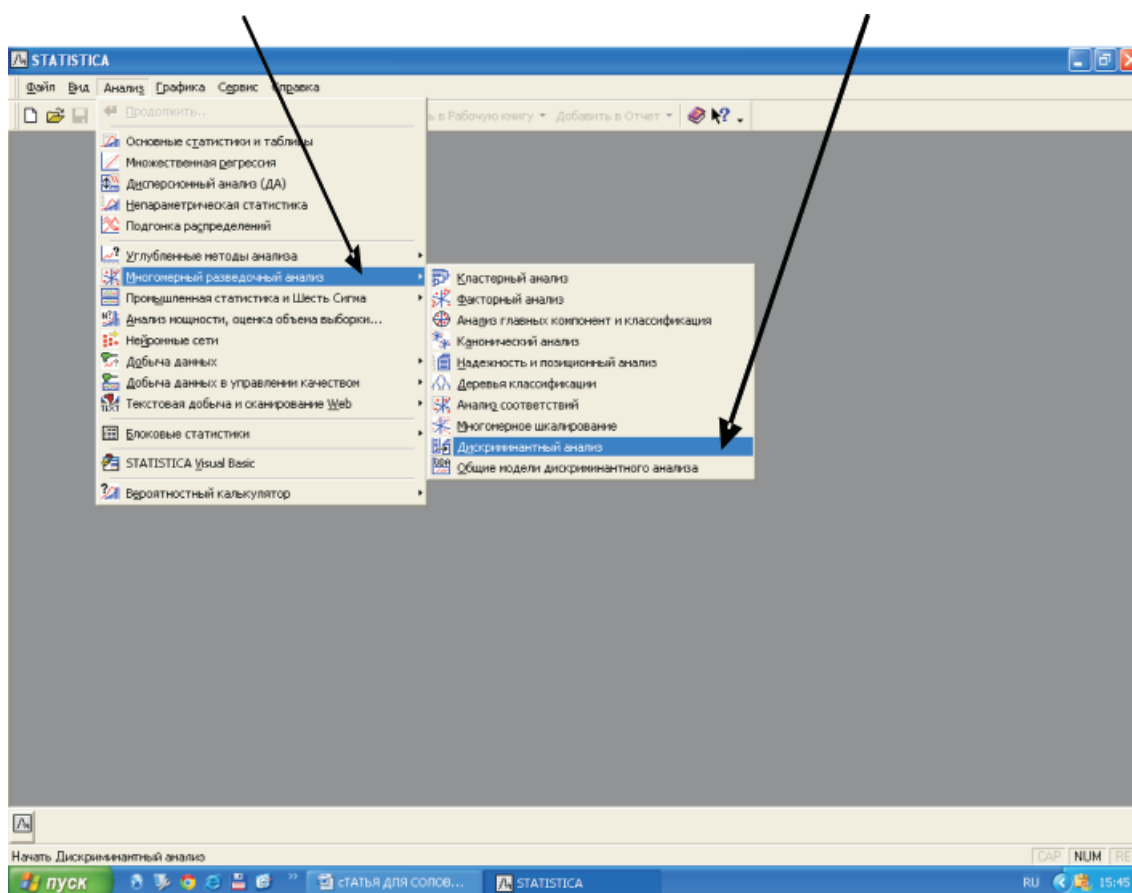


Рис. 1. Окно Анализ программы «Statistica» v.6

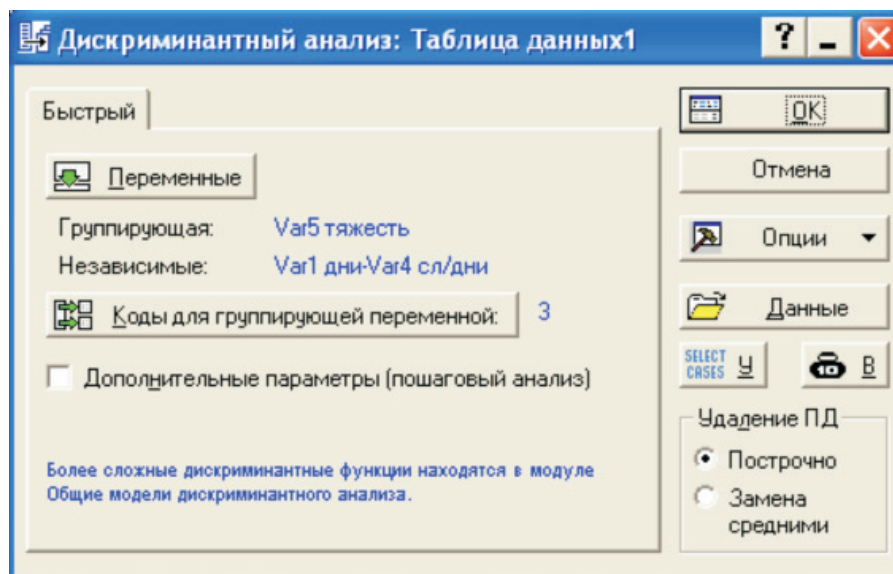


Рис. 2. Окно Дискриминантный анализ программы «Statistica» v.6

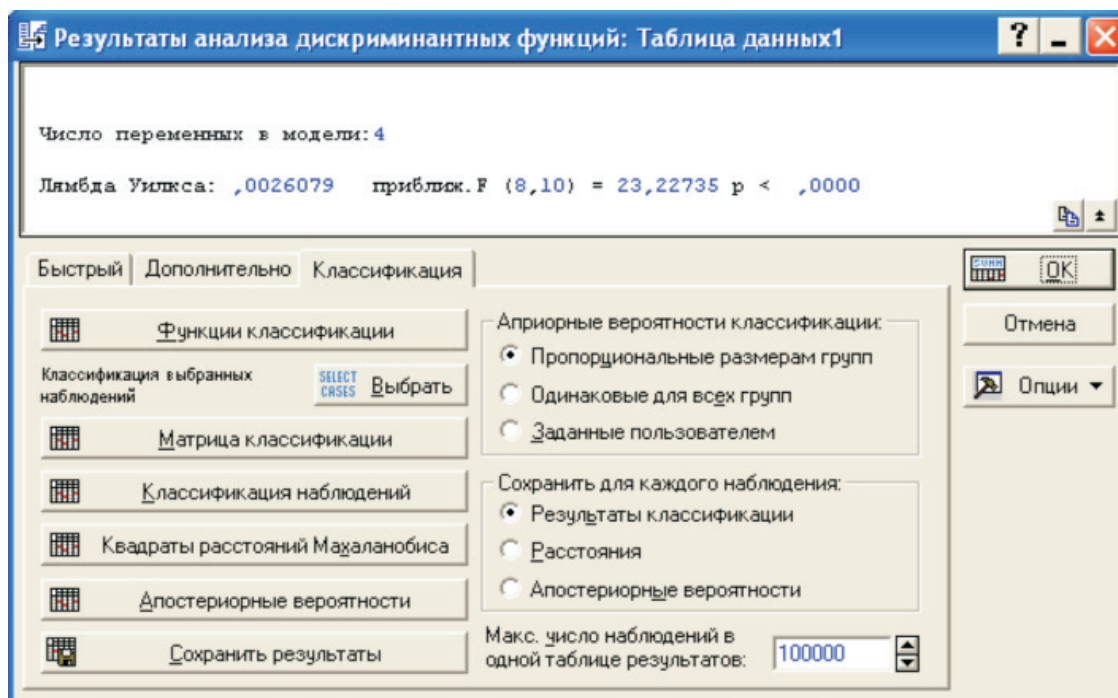


Рис. 3. Окно Результаты анализа

Для подтверждения и закрепления факта дискриминации определим еще несколько показателей. На первом месте – **Расстояние Махаланобиса**, которое является мерой близости отдельно взятых наблюдений и центром каждой совокупности, из включенных в процесс дискриминации. Чем ближе наблюдение к центроиду конкретной совокупности, тем в большей степени можно быть уверенным, что наблюдение извлечено имен-

но из неё. **Расстояние Махаланобиса** может быть рассчитано при нажатии на кнопку **Квадраты расстояния Махаланобиса** во вкладке **Классификация**. Дифференциация случаев травмирования по этому признаку отражена в табл. 3 (цветом выделены статистически значимые показатели).

Кроме **Расстояния Махаланобиса** можно вычислить еще и условную (или апостериорную) вероятность принадлежности

наблюдения к определенной совокупности. Её условность в том, что она зависит от знания исследователем значений переменных в модели. Этот показатель получают, нажав на кнопку **Апостериорные вероятности**.

В данном примере точность классификации очень высока, даже с учетом того, что это апостериорная классификация. К слову сказать, такая точность редко достигается и редко, когда нужна.

Таблица 2

Квадраты расстояний Махаланобиса

Квадраты расстояний Махаланобиса (Таблица данных 1)				
№ п/п	Тяжесть травмы	Легкая	Средней тяжести	Тяжелая
1	Легкая	1,118	1011,483	688,3675
2	Легкая	3,017	929,539	620,7845
3	Средней тяжести	1042,041	0,971	36,0575
4	Тяжелая	741,390	23,638	4,0505
5	Средней тяжести	1048,528	3,473	43,2491
6	Средней тяжести	1073,374	2,437	41,9351
7	Легкая	3,473	1048,528	721,5672
8	Средней тяжести	934,432	4,560	21,1041
9	Тяжелая	678,040	45,932	3,5372
10	Тяжелая	682,858	37,428	1,2326
11	Легкая	4,130	1109,123	772,3102

Для проверки работоспособности представленной модели с учетом вероятностей

в исходную табл. 1 введем переменные под № 12, 13, 14 с их значениями, табл. 3

Таблица 3

Проверка работоспособности методики анализа

№ п/п	Var1 Дни	Var2 Случаи	Var3 Стоимость	Var4 Случаи/дни	Var5 Тяжесть травмы
1	50	3	1,2	0,06	Легкая
2	50	3	1,4	0,06	Легкая
3	64	2	5,6	0,05	Средней тяжести
4	65	2	4,8	0,05	Тяжелая
5	67	3	5,6	0,04	Средней тяжести
6	63	3	5,7	0,05	Средней тяжести
7	46	4	1,4	0,06	Легкая
8	69	3	5,1	0,04	Средней тяжести
9	62	2	4,5	0,03	Тяжелая
10	59	3	4,8	0,05	Тяжелая
11	45	4	1,3	0,08	Легкая
12	44	4	1	0,09	
13	43	5	1,1	0,13	
14	67	2	6	0,03	

При повторении анализа машина мгновенно классифицирует травмы по тяжести, отнеся 12 и 13 случаи к легким, а 14 – к среднетяжелым травмам, табл. 4. Примечательно, что классификация на-

блюдений по вероятностным признакам оказалась гораздо показательней расчета квадратов **Расстояний Махаланобиса**: дифференциация в данном случае равна 1,0 или 100 %.

Таблица 4

Апостериорные вероятности травмирования

Апостериорные вероятности				
№ п/п	Тяжесть травмы	Легкая	Тяжелая	Средней тяжести
1	Легкая	1,000000	0,000000	0,000000
2	Легкая	1,000000	0,000000	0,000000
3	Средней тяжести	0,000000	0,000000	1,000000
4	Тяжелая	0,000000	0,999996	0,000004
5	Средней тяжести	0,000000	0,000000	1,000000
6	Средней тяжести	0,000000	0,000000	1,000000
7	Легкая	1,000000	0,000000	0,000000
8	Средней тяжести	0,000000	0,000001	0,999999
9	Тяжелая	0,000000	1,000000	0,000000
10	Тяжелая	0,000000	1,000000	0,000000
11	Легкая	1,000000	0,000000	0,000000
12	---	1,000000	0,000000	0,000000
13	---	1,000000	0,000000	0,000000
14	---	0,000000	0,000000	1,000000

Выводы

Хотя данный пример нами сознательно упрощен, тем не менее, хорошо иллюстрирует основную идею дискриминации. Для «перестраховки» в ответственных случаях следует проводить дискриминацию в два этапа: *сначала построить функции классификации и только потом проводить оценку их качества.*

При использовании данного вида анализа необходимо учитывать несколько ограничений: нормальность и линейность эмпирического распределения, однородность дисперсий и ковариаций сравниваемых совокупностей. Однако, как показала наша практика, методика достаточно «терпима» к отклонениям от этих условностей.