

## ДЗ - компьютерная обработка

1.1. Вводные данные:

- группирующая переменная - "оценка";
- независимые переменные - с "и1" до "н" ("и1"- "и11", "э\_и", "н");

1.2. Провести дискриминантный анализ по данным файла "class"

(размещен в 2-х форматах: .sav (IBM SPSS) и .xlsx (MS Excel);

1.3. Записать выражение для дискриминантной функции;

<b>Коэффициенты канонической дискриминантной функции</b>	
	Функция 1
осведомленность	,024
скрытые фигуры	,146
пропущенные слова	-,010
счет в уме	,304
понятливость	,217
исключение изображений	-,014
анalogии	,166
числовые ряды	,107
умозаключения	-,265
геометрическое сложение	,084
заучивание слов	,161
экстраверсия	,086
нейротизм	-,095
(Константа)	-9,132
Нестандартизованные коэффициенты	

Дискриминантная функция:  $f(x) = -9,132 + 0,24x_1 + 0,146x_2 - 0,1x_3 + 0,304x_4 + 0,217x_5 - 0,14x_6 + 0,166x_7 + 0,107x_8 - 0,265x_9 + 0,84x_{10} + 0,161x_{11} + 0,86x_{12} - 0,95x_{13}$

Где  $x_1 \dots x_{13}$  - это все независимые переменные от осведомленности до нейротизма соответственно.

1.4. Оценить значимость дискриминантной функции (по коэффициенту Уилкса);

Лямбда Уилкса				
Критерий для функций	Лямбда Уилкса	Хи-квадрат	ст.св.	знач.
1	,488	26,870	13	,013

В нашем примере лямбда Уилкса равна 0,488, что является не таким плохим результатом, однако было бы лучше, если бы она была ближе к 0. Функция значима, так как уровень значимости меньше 0,05.

1.5. Определить относительный вклад каждой переменной в формирование классов;

Коэффициенты стандартизированной канонической дискриминантной функции	
	Функция 1
осведомленность	,051
скрытые фигуры	,371
пропущенные слова	-,026
счет в уме	,537
понятливость	,393
исключение изображений	-,033
анalogии	,449
числовые ряды	,242
умозаключения	-,776
геометрическое сложение	,206
заучивание слов	,347
экстраверсия	,287
нейротизм	-,376

Относительный вклад каждой переменной в различие двух сформированных классов можно оценить с помощью коэффициентов стандартизированной канонической дискриминантной функции. Чем больше коэффициент по модулю, тем больше влияние на формирование классов оказывает переменная. Таким образом умозаключения, счет в уме и аналогии имеют наибольший вклад, а осведомленность и исключение изображений наименьший.

1.6. Определить, чему равны средние значения дискриминантной функции по группам;

Функции в центроидах групп	
оценка успешности	Функция 1
низкая	-1,001
высокая	1,001
Нестандартизованные канонические дискриминантные функции, вычисленные в групповых средних	

Расстояние между центроидами – среднее значение дискриминантной функции в исследуемых группах. В группе с низкой оценкой успешности (1) среднее значение дискриминантной функции равно = -1,001. В группе с высокой оценкой успешности (2) среднее значение дискриминантной функции равно = 1,001.

1.7. Указать, к каким группам были отнесены классифицируемые объекты и вероятности, с которыми объекты входят в эти группы;

Статистика по наблюдениям						
			Наивысшая группа			
			P(D>d   G=g)			
	Номер наблюдения	Фактическая группа	Предсказанная группа	PM	ст.св.	P(G=g   D=d)
Исходный	1	1	1	,681	1	,765
	2	2	2	,927	1	,861
	3	1	1	,768	1	,804
	4	2	2	,527	1	,963
	5	2	2	,478	1	,968
	6	2	2	,454	1	,624
	7	1	1	,263	1	,986
	8	2	2	,928	1	,861
	9	1	1	,460	1	,629
	10	2	2	,072	1	,996
	11	1	1	,354	1	,979
	12	2	2	,592	1	,956
	13	2	2	,668	1	,946
	14	2	2	,720	1	,784
	15	2	1**	,651	1	,750
	16	1	1	,466	1	,633
	17	1	1	,831	1	,919
	18	1	1	,981	1	,876
	19	2	2	,752	1	,933
	20	1	2**	,932	1	,898
	21	2	2	,564	1	,959
	22	1	1	,077	1	,996
	23	1	1	,787	1	,812
	24	2	2	,749	1	,796
	25	2	2	,546	1	,689

26	1	1	,897	1	,906
27	1	1	,808	1	,820
28	1	1	,907	1	,854
29	2	2	,124	1	,994
30	1	1	,593	1	,718
31	2	2	,342	1	,980
32	2	1**	,137	1	,993
33	2	2	,605	1	,954
34	1	2**	,320	1	,504
35	2	1**	,359	1	,542
36	2	2	,421	1	,597
37	2	2	,326	1	,981
38	1	1	,615	1	,953
39	1	1	,984	1	,877
40	2	2	,871	1	,843
41	1	1	,161	1	,992
42	1	1	,226	1	,988
43	1	1	,659	1	,754
44	2	2	,747	1	,934
45	1	1	,800	1	,925
46	1	1	,543	1	,687
47	не сгруппирован о	2	,837	1	,831
48	не сгруппирован о	2	,105	1	,995
49	не сгруппирован о	1	,320	1	,503
50	не сгруппирован о	1	,962	1	,871
51	не сгруппирован о	1	,501	1	,659
52	не сгруппирован о	2	,757	1	,800
53	не сгруппирован о	2	,686	1	,767
54	не сгруппирован о	1	,467	1	,633
55	не сгруппирован о	1	,919	1	,858
56	не сгруппирован о	1	,424	1	,973
**. Ошибочно классифицированное наблюдение					

У большинства классифицируемых объектов фактическая и предсказанные группы совпали. Исключением стали наблюдения 15, 20, 32, 34 и 35. Не сгруппированные наблюдения 49-51 и 54-56 попали в группу 1, а объекты 47, 48, 52 и 53 в группу 2. Вероятности, с которыми наблюдения входят в предсказанные группы можно увидеть в столбце с названием Р.

1.8. На основании анализа таблицы «Wilks' Lambda» проверьте значимость различий средних значений дискриминантной функции в двух группах;

**Критерии равенства групповых средних**

	Лямбда Уилкса	F	ст.св.1	ст.св.2	знач.
осведомленность	,901	4,859	1	44	,033
скрытые фигуры	,850	7,746	1	44	,008
пропущенные слова	,931	3,271	1	44	,077
счет в уме	,806	10,563	1	44	,002
понятливость	,844	8,127	1	44	,007
исключение изображений	,935	3,064	1	44	,087
анalogии	,866	6,811	1	44	,012
числовые ряды	,915	4,075	1	44	,050
умозаключения	,996	,163	1	44	,689
геометрическое сложение	,910	4,375	1	44	,042
заучивание слов	,908	4,479	1	44	,040
экстраверсия	,986	,627	1	44	,433
нейротизм	,997	,140	1	44	,710

Переменные пропущенные слова, исключение изображений, умозаключения, экстраверсия и нейротизм можно убрать из анализа, т.к. значимость их различия с другими значениями их группы больше 0,05. Остальные переменные имеют значимость меньше или равно 0,05 и следовательно средние двух групп значимо различаются, т.е. доказано наличие дискриминирующих особенностей этих переменных.

1.9. Оценить качество дискриминантного анализа (на основании результатов таблицы Eigenvalue);

**Собственные значения**

Функция	Собственное значение	% дисперсии	Суммарный %	Каноническая корреляция
1	1,047 <sup>а</sup>	100,0	100,0	,715

а. Для анализа использовались первые 1 из канонических дискриминантных функций.

Собственное значение - отношение межгрупповой дисперсии к внутригрупповой дисперсии выборочных значений дискриминантной

функции. Чем больше собственное значение, тем лучше подобрана дискриминантная функция.

В нашем случае собственное значение равно 1,047, что является относительно небольшим значением. А значит дискриминантная функция могла бы быть подобрана лучше.

Каноническая корреляция характеризует качество достоверности дискриминации.

В нашем случае достоверность дискриминации равно 0,715, что опять же является достаточно средним показателем и с другой функцией дискриминация могла бы быть лучше.

1.10. Оценить целесообразность проведения дискриминантного анализа по Вашим данным.

Я считаю, что по предоставленным данным нецелесообразно проводить дискриминантный анализ. Дискриминантная функция оказалась не очень хорошо подобранной, а больше 30% переменных не значимыми для анализа. Возможно, что повторное проведение анализа после удаления незначимых переменных покажет более качественные результаты.