

ДИСПЕРСИОННЫЙ АНАЛИЗ

Дисперсионный анализ – статистический метод, позволяющий анализировать влияние различных факторов на исследуемую переменную. *Однофакторный дисперсионный анализ* выявляет влияние некоторого фактора, под действием которого общая выборка делится на t частных выборок (групп). Рассмотрим пример.

ПРИМЕР 1. В институте 6 факультетов. Ставится задача, на уровне значимости $\alpha=0,05$ определить, различаются или нет уровни подготовки по иностранным языкам на факультетах, то есть влияет ли фактор «Факультет» на уровень подготовки. Для этой цели были протестированы студенты всех факультетов по иностранному языку, результаты тестирования (баллы) приведены в таблице:

Факультет 1	43	81	75	94	64	51	34	84	39	69	59	83
Факультет 2	63	82	61	78	94	58	84	79	43	96	89	
Факультет 3	99	50	54	92	33	90	56	46	75	55	35	85
Факультет 4	78	57	52	31	69	51	48	36	94	75		
Факультет 5	37	74	33	66	33	79	81	75	33	87	83	92
Факультет 6	80	58	70	93	56	82	35	53	99	59	48	

Вводим эту таблицу вместе с подписями в Excel в ячейки от A1 до шестой строки и столбца M. Затем вызываем надстройку «Анализ данных». Она

В окне «Анализ данных» нужно выбрать пункт «Однофакторный дисперсионный анализ». В открывшемся окне в поле «Входной интервал» делаем ссылку на диапазон A1-M6, группирование – «по строкам», ставим галочку напротив «Метки в первом столбце», альфа – 0,05, в разделе «Параметры вывода» ставим точку рядом с «Выходной интервал» и в поле рядом делаем ссылку на ячейку, с которой начнется вывод данных, например с A9, нажимаем ОК.

На рабочем листе появиться таблица с результатами однофакторного дисперсионного анализа. Она состоит из двух частей. В первой «Итоги» приведены основные статистические показатели по выборкам: их объем, сумма элементов, среднее и дисперсия. Во второй «Дисперсионный анализ» - непосредственно расчет критерия. В строках «Между группами» и «Внутри групп» приведены остаточные суммы (SS), степени свободы (df) и дисперсии (MS). Далее приведена F-статистика (большая дисперсия на меньшую), критическое значение уровня значимости (P -значение) и F-критическое. Если F-статистика больше F-критического, или критическое значение уровня значимости меньше заданного α , то средние в выборках (группах) различаются и фактор влияет на показатель. У нас этого не наблюдается, следовательно, уровень подготовки на всех факультетах можно считать одинаковым.

Задание 1. В школе 6 пятых классов. Психологу ставится задача, определить, одинаковый ли средний уровень ситуативной тревожности в классах. Для этого были проведены тесты тревожности среди школьников, результаты которых следующие:

5 «А» класс	54	32	24	33	49	24	26	32	50	52	33	32	36
5 «Б» класс	64	44	33	50	35	40	66	38	46	42	44	68	43
5 «В» класс	30	39	48	30	31	21	39	19	23	40	43	40	
5 «Г» класс	36	24	46	19	47	47	34	40	32	22	17	20	
5 «Д» класс	51	44	38	55	28	28	29	41	39	25	61		
5 «Е» класс	27	56	49	37	21	28	47	30	44	48	28	25	21

Проверить на уровне значимости $\alpha=0,02$ предположение, что средняя ситуативная тревожность в классах не различается.

Задание 2. Студентов разделили на 8 групп. При этом предполагается, что средний уровень знаний в группах равный. Результаты тестирования по дисциплине представлены в таблице

Группа	Обучающийся														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Группа 1	54	32	24	33	49	24	26	32	50	52	33	32	36	54	32
Группа 2	64	44	33	50	35	40	66	38	46	42	44	68	43	64	
Группа 3	30	39	48	30	31	21	39	19	23	40	43	40	44	30	39
Группа 4	36	24	46	19	47	47	34	40	32	22	17	20	24		
Группа 5	51	44	38	55	28	28	29	41	39	25	61	36	48	51	
Группа 6	27	56	49	37	21	28	47	30	44	48	28	25	21	27	56
Группа 7	54	32	24	33	49	24	26	32	50	52	33	32			
Группа 8	64	44	33	50	35	40	66	38	46	42	44	68	43	64	

Проверить на уровне значимости $\alpha=0,05$ предположение, что средняя подготовленность обучающихся в группах не различается.

Рассмотрим теперь пример *двухфакторного дисперсионного анализа*. В этом случае методика расчета та же, но общая выборка распадается на группы под влиянием двух факторов, влияние одного из них сгруппируем в строках, а второго – в столбцах. Рассмотрим решение на примере.

ПРИМЕР 2. Предположим теперь, что на качество освоения иностранного языка может влиять не только факультет (которых шесть), но и сам язык (которых три: английский, немецкий и французский). Тогда с каждого факультета были взяты по 5 студентов, изучающих разные языки. Выборки результатов тестов (баллы) приведены в таблице.

ЯЗЫК	Факультет 1	Факультет 2	Факультет 3	Факультет 4	Факультет 5	Факультет 6
Англий-ский	54	56	50	48	89	85
	33	97	98	82	34	72
	78	84	54	34	46	78
	93	65	67	44	50	40
	85	86	72	31	91	49

ЯЗЫК	Факультет 1	Факультет 2	Факультет 3	Факультет 4	Факультет 5	Факультет 6
Немец- кий	81	89	82	67	77	35
	44	39	85	35	70	73
	91	39	55	38	56	47
	72	43	69	69	49	82
	89	97	49	78	67	68
Фран- цузский	92	74	56	86	57	71
	91	83	88	85	73	89
	84	97	89	65	80	85
	65	55	69	54	60	75
	40	89	95	88	83	49

Нужно проверить следующие гипотезы (на уровне значимости $\alpha=0,05$):

1. Влияет ли факультет на уровень подготовки.
2. Влияет ли язык на уровень подготовки.
3. Влияют ли друг на друга (коррелируют) язык и факультет.

Переходим на новый рабочий лист. Вводим данные из таблицы вместе с подписями в ячейки A1-G16. В первом столбце группировать ячейки не нужно, просто введите подписи «английский» в A2, «немецкий» в A7 и «французский» в A12. Вызывает надстройку «Анализ данных» и в ней – «Двухфакторный дисперсионный анализ с повторениями». В открывшемся окне в поле «Входной интервал» делаем ссылку на диапазон A1-G16, в поле «Число строк для выборки» вводим 5, альфа – 0,05, в разделе «Параметры вывода» ставим точку рядом с «Выходной интервал» и в поле рядом делаем ссылку на ячейку, с которой начнется вывод данных, например с A18, нажимаем ОК.

На рабочем листе появиться область с результатами двухфакторного дисперсионного анализа. Она состоит из нескольких таблиц, в которых приведены основные статистические показатели для факультетов и языков. В последней таблице «Дисперсионный анализ» приведены результаты расчета критерия. Первая строка «Выборка» отображает результаты по языкам. Видно, что F-статистика больше, чем F-критическое и критический уровень значимости P -значение меньше заданного 0,05. Следовательно, средние результаты теста для языков значимо различаются и фактор «Язык» влияет на качество освоения языков.

Во второй строке «Столбцы» отображаются результаты по факультетам. Видно, что F-статистика меньше, чем F-критическое и критический уровень значимости P -значение больше заданного 0,05. Следовательно, средние результаты теста для факультетов не различаются и фактор «Факультет» не влияет на качество освоения языков.

В третьей строке «Взаимодействие» отображаются результаты выявления зависимости факторов «Язык» и «Факультет» друг на друга. Видно, что F-статистика меньше, чем F-критическое и критический уровень значимости P -значение больше заданного 0,05. Следовательно, факторы «Язык» и «Факультет» не влияют по качеству освоения языков друг на друга.

Задание 3. Ставится задача, определить, влияет ли тип темперамента и образование на производительность труда рабочих предприятия. Для этого были отобраны по 4 рабочих каждого типа темперамента образования и измерены производительности труда. Проверить на уровне значимости $\alpha=0,02$ влияние темперамента и образование на производительность труда и друг на друга.

Образова- ние	Высшее	Средне- профес.	Среднее общее	Среднее полное	Образова- ние	Высшее	Средне- профес.	Среднее общее	Среднее полное
Тем- перамент Холерик	386	498	400	347	Тем- перамент Флегматик	387	383	304	335
	240	334	218	456		263	492	231	422
	209	389	230	258		338	295	237	262
	403	327	247	420		397	315	351	452
Сангвиник	371	251	368	270	Меланхо- лик	366	370	460	359
	339	352	469	327		261	231	421	229
	298	258	260	422		357	343	283	345
	322	299	440	372		275	241	372	206

Задание 4. Менеджер распределяет группы из 5 сотрудников по 6 районам с целью выполнения 4 видов работ. Имеются данные о предыдущих показателях эффективности (процент прибыли от работы) этих сотрудников в районах по каждой работе. Они указаны в таблице. Проверить на уровне значимости $\alpha=0,05$ влияние района и вида работ на эффективность, а также влияние друг на друга района и эффективности.

Работа	Работа1	Работа2	Работа3	Работа4	Работа	Работа1	Работа2	Работа3	Работа4
Район Район 1	17,7	27,3	24,9	30,3	Район Район 4	36,7	23,2	24,5	25,7
	31,2	28,7	25,4	33,9		27,0	26,1	33,2	21,8
	25,0	16,1	21,5	26,9		22,8	24,4	31,0	27,5
	19,3	23,5	22,1	27,2		26,0	18,9	15,4	29,0
	19,5	29,5	17,6	21,3		19,0	25,3	22,4	32,0
Район 2	17,9	16,7	19,9	21,3	Район 5	12,8	16,0	16,2	29,9
	14,3	18,3	12,8	21,0		10,9	4,0	23,6	11,5
	13,4	23,1	16,1	15,9		21,5	30,4	11,8	15,4
	17,6	15,2	20,5	18,0		27,8	24,3	16,0	15,9
	14,3	17,6	15,9	20,7		19,4	21,7	14,2	24,4
Район 3	34,6	25,0	27,0	20,6	Район 6	29,6	26,7	27,5	33,1
	34,5	27,5	10,3	31,1		28,1	23,9	29,5	31,1
	27,1	21,6	21,4	13,2		32,1	30,8	28,4	23,5
	26,7	18,2	32,8	20,3		29,0	36,2	24,6	35,3
	18,2	37,9	15,6	26,9		31,1	30,9	32,4	27,0