

Дисперсионный анализ.

Сущность его заключается в установлении роли отдельных факторов в изменчивости того или иного признака. Дело в том, что влияние тех или иных факторов на изучаемый признак не может быть выделено в чистом виде, различные опыты дают несколько неодинаковые результаты. Объясняется это тем, что на них влияют многочисленные случайные обстоятельства, многие другие факторы, несколько меняющиеся от опыта к опыту и не поддающиеся контролю. Вот почему возникает важная задача разложения общей изменчивости признака на составные части, с одной стороны, определяемые изучаемым конкретным фактором, а с другой – вызываемые случайными неконтролируемыми причинами.

Дисперсионный анализ позволяет оценивать значимость влияния отдельных факторов, а так же их относительную роль в общей изменчивости признака.

В фактическом отклонении варианты от средней генеральной совокупности фигурируют два компонента:

- та часть отклонения, которая зависит именно от данного фактора – **A**;
- остаточная часть, независимая от данного фактора – **e**.

В таком случае можно сравнивать **A** и **e**.

При достоверном влиянии изучаемого фактора значение **A** будет превышать значение **e**.

По степени превышения **A** над **e** можно судить о том, насколько достоверно влияние данного фактора.

Разберём простейшую схему, когда анализируется влияние одного фактора, могущего принимать разные градации, или количественные уровни: **1,2,3... i ...a**.

Отдельные наблюдения (варианты) разбиваются на группы согласно этим градациям фактора. Количество наблюдений в одном уровне: **1,2,3 ...j ...n**.

Общее количество наблюдений равно **N=a·n**.

Распределение вариантов при различии по одному фактору представлено в таблице:

n-число наблюдений в каждой группе;
a-количество групп (уровней фактора A);
N-количество всех вариантов (N= n·a);
i-разные уровни;
j-разные наблюдения.

A	Отдельные варианты (наблюдения)						$\sum X_i = T$ i	T_i^2
	1	2	...	j	...	N		
1	X ₁₁	X ₁₂		X _{1j}		X _{1n}	$\sum X_1 = T_1$	T_1^2
2	X ₂₁	X ₂₂		X _{2j}		X _{2n}	$\sum X_2 = T_2$	T_2^2
:								
I	X _{i1}	X _{i2}		X _{ij}		X _{in}	$\sum X_i = T$ i	T_i^2

:								
a	X_{a1}	X_{a2}		X_{aj}		X_{an}	$\sum X_{a=} = T_a$	T_a^2
							$\sum T_i = T$	$\sum T_i^2 =$
							$T^2 =$	

Найдём сумму квадратов, составив дополнительную таблицу ($\sum X_{ij}^2$).

X_{11}^2	X_{12}^2	...	X_{1j}^2	...	X_{1n}^2
X_{21}^2	X_{22}^2	...	X_{2j}^2	...	X_{2n}^2
...	
X_{i1}^2	X_{i2}^2	...	X_{ij}^2	...	X_{in}^2
...	
X_{a1}^2	X_{a2}^2	...	X_{aj}^2	...	X_{an}^2

Степени свободы:

1. Для общей дисперсии $k_0 = N - 1$, где $N = a \cdot n$
2. Для факториальной дисперсии $K_A = a - 1$
3. Для остаточной дисперсии $K_e = N - a$

Формулы для вычисления дисперсий.

Общая дисперсия $\sigma_0^2 = \frac{1}{N-1} (\sum X_{ij}^2 - \frac{T^2}{N})$

Факториальная дисперсия $\sigma_A^2 = \frac{1}{a-1} (\sum \frac{T_i^2}{n} - \frac{T^2}{N})$

Остаточная дисперсия $\sigma_e^2 = \frac{1}{N-a} (\sum X_{ij}^2 - \sum \frac{T_i^2}{n})$

Нулевая гипотеза.

При дисперсионном анализе следует исходить из первоначально принимаемой нулевой гипотезы, что данный фактор А не влияет на изменчивость данного признака.

Если правильна нулевая гипотеза, то σ_A^2 должна быть равна нулю, т.е. вся вариация сводится только к случайной. Для того, чтобы отбросить нулевую гипотезу, нужно доказать, что σ_A^2 – достоверно (т. е. с вероятностью, не меньше, чем 0.95, или с $\alpha = 0.05$) отличается от нуля. Достоверность значения σ_A^2 может быть установлена путём деления на ошибку, т. е. на σ_e^2 :

$$F = \frac{\sigma_A^2}{\sigma_e^2}$$

Количественная оценка влияния отдельных факторов.

Наряду с доказательством влияния того или иного фактора на результативный признак, часто возникает необходимость установления меры этого влияния и его доли в сумме влияния всех факторов. Доля влияния фактора А равна:

$$P_A = \frac{\sigma_A^2 - \sigma_e^2}{\sigma_A^2 + \sigma_e^2 (n-1)}$$

Задача:

Получены следующие данные о содержании каратиноидов (в мг/кв.дм) в листьях канатика в разное время суток. Влияет ли время суток на содержание каратиноидов?

Часы суток	1	2	3	4
15	1,41	0,95	1,00	0,93
18	1,17	1,10	0,84	1,01
21	1,38	1,38	0,91	1,36
24	0,62	0,48	0,43	0,62
6	0,74	0,41	0,41	0,43
9	0,76	0,59	0,74	0,46
12	0,64	1,02	1,04	0,98

Решение:

Данные заносят в таблицу и делают расчёты.

A	Отдельные наблюдения. n				$T_i = \sum X_i$	T_i^2
	1	2	3	4		
1	1,41	0,95	1,00	0,93	4,29	18,4041
2	1,17	1,10	0,84	1,01	4,12	16,9744
3	1,38	1,38	0,91	1,36	5,03	25,3009
4	0,62	0,48	0,43	0,62	2,15	4,6225
5	0,74	0,41	0,41	0,43	1,96	3,9601
6	0,76	0,59	0,74	0,46	2,55	6,5025
7	0,64	1,02	1,04	0,98	3,68	13,5421

$$T=23,81 \quad \sum T_i^2=89,3$$

$$T^2=566,9$$

Схема решения задачи на дисперсионный анализ:

1. Суммируют данные задачи по каждому уровню фактора A (T_i).
2. Находят общую сумму ($\sum T_i$) по всем уровням и получают значение (T).
3. Возводят в квадрат общую сумму ($\sum T_i$) и получают значение T^2
4. Возводят полученные суммы (T_i) в квадрат и получают значения (T_i^2) по каждому уровню.
5. Находят общую сумму ($\sum T_i^2$)
6. Находят сумму квадратов данных задачи, составив дополнительную таблицу.

			$\sum X_{ij}^2=$	22,9363
1,9881	0,9025	1,00	0,8649	4,7555

1,3689	1,21	0,7056	1,0201	4,3046
1,9044	1,9044	0,8281	1,8496	6,4865
0,3844	0,2304	0,1849	0,3844	1,1841
0,5446	0,1681	0,1681	0,1849	1,6163
0,5776	0,3481	0,5476	0,2116	1,6849
0,4096	1,0404	1,0816	0,9604	3,492

7. Определяют число степеней свободы:

1. Для общей дисперсии $k_0 = N - 1 = 28 - 1 = 27$
2. Для факториальной дисперсии $K_A = a - 1 = 7 - 1 = 6$
3. Для остаточной дисперсии $K_e = N - a = 28 - 7 = 21$

8. Рассчитывают дисперсии:

Общая дисперсия : $\sigma_0^2 = \frac{1}{N-1} \left(\sum X_{ij}^2 - \frac{T^2}{N} \right)$

$$\sigma_0^2 = \frac{1}{27} \cdot \left(22,7316 - \frac{566,9161}{28} \right) = 0,092$$

Факториальная дисперсия $\sigma_A^2 = \frac{1}{a-1} \left(\sum \frac{T_i^2}{n} - \frac{T^2}{N} \right)$

$$\sigma_A^2 = \frac{1}{6} \cdot \left(\frac{89,3069}{4} - 20,2470 \right) = \frac{2,0797}{6} = 0,3466$$

Остаточная дисперсия $\sigma_e^2 = \frac{1}{N-a} \left(\sum X_{ij}^2 - \sum \frac{T_i^2}{n} \right)$

$$\sigma_e^2 = \frac{1}{21} \cdot \left(22,7316 - \frac{89,3063}{4} \right) = 0,0193$$

9. Находят фактическое значение F-критерия Фишера:

$$F_\phi = 0,3466 / 0,0193 = 18$$

10. Находят стандартное значение F-критерия Фишера (по таблице №4) $F_{st} = 3,84$ при уровне значимости 0,01 и числа степеней свободы для большей дисперсии ($K_A = 7 - 1 = 6$) и меньшей дисперсии ($K_e = 28 - 7 = 21$).

11. Сравнивают фактически найденное значение F-критерия Фишера со стандартным значением и делают вывод.

$F_\phi \geq F_{st}$, нулевую гипотезу отвергают на 1% уровне значимости.

Вывод: С вероятностью более 0,99 можно заключить, что время суток влияет на содержание каратиноидов в листьях канатика.

12. Определяют степень влияния фактора А по формуле: $P = \frac{\sigma_A^2 - \sigma_e^2}{\sigma_A^2 + \sigma_e^2(n-1)}$

$$P = \frac{0,3466 - 0,0193}{0,3466 + 0,0193(4-1)} = 0,81$$

$P = 0,81$ – Это означает, что примерно около 81% от общего варьирования содержания каратиноидов в листьях канатика обусловлено влиянием времени суток

Ответ:

1. С вероятностью более 0,99 можно заключить, что время суток влияет на содержание каротиноидов в листьях канатика.
2. Сила влияния фактора А на результирующий признак равна $P=0,81$.

Двухфакторный дисперсионный анализ

Пример:

С помощью дисперсионного анализа. Выявить влияние реагентов А и В на синтез лекарственного препарата.

Уровни А	Уровни В			
	$B_1 = 1$	$B_2 = 2$	$B_3 = 3$	$B_4 = 4$
3	7,5	3,0	4,0	3,5
6	3,5	2,5	3,5	2,0
9	6,5	5,5	4,5	6,0
11	7,5	7,0	8,5	7,0

Решение:

Данные занесем в таблицу и сделаем расчеты:

В \ А	1	2	3	4	T_j	T_j^2
1	7,5	3,0	4,0	3,5	15	225
2	3,5	2,5	3,5	2,0	11,5	132,25
3	6,5	5,5	4,5	6,0	22,5	506,25
4	7,5	7,0	8,5	7,0	30	900
$a=4$						
T_i	22	18	20,5	18,5	$\sum T_i=79$	$\sum T_j^2=1763,5$
					$T^2=6241$	
T_i^2	484	324	420,25	342,25	$\sum T_i^2=1570,5$	

Найдем сумму квадратов, составив дополнительную таблицу: $\sum X_{ij}^2=483,5$

56,25	9	16	12,25
12,25	6,25	12,25	4
42,25	30,25	20,25	36
56,25	49	72,25	49

$$\sigma_A^2 = \frac{1}{n(a-1)} \cdot \left(\sum T_{ji}^2 - \frac{T^2}{a} \right) = \frac{1}{4 \cdot 3} \left(1763,5 - \frac{6241}{4} \right) = 16,9$$

$$\sigma_B^2 = \frac{1}{a(n-1)} \cdot \left(\sum T_i^2 - \frac{T^2}{n} \right) = \frac{1}{4 \cdot 3} \left(1570,5 - \frac{6241}{4} \right) = 0,85$$

$$\sigma_e^2 = \frac{1}{(n-1)(a-1)} \cdot \left(\sum X_{ij}^2 - \frac{\sum T_i^2}{a} - \frac{\sum T_j^2}{n} + \frac{T^2}{a \cdot n} \right) = 0,45$$

$$F_{\phi}^A = \frac{16,9}{0,45} = 37,5 \quad F_{\phi}^B = \frac{0,85}{0,45} = 1,89$$

Степени свободы $R_A=m-1=3$; $R_B=n-1=3$; $R_e=(n-1)(m-1)=9 \rightarrow F_{st}=3,86$ ($\alpha=5\%$).

Вывод: 1. $F_{\phi}^A \geq F_{st} \rightarrow$ гипотеза H_0 отвергается, реагент А влияет на выход препарата.

$$\text{Степень влияния фактора } P = \frac{\sigma_A^2 - \sigma_e^2}{\sigma_A^2 + \sigma_e^2(n-1)} = 0,90$$

2. $F_{\phi}^B < F_{st} \rightarrow$ гипотеза H_0 сохраняется, реагент В не влияет на выход препарата.

Задачи для самостоятельного решения на дисперсионный анализ.

1 Исследовали интенсивность распада белка в организме обожженных собак при введении нормальной и иммунной сыворотки. Используя метод дисперсионного анализа, определить достоверность влияния иммунной сыворотки на снижение интенсивности распада белка у обожженных собак.

Вид сыворотки	№ испытания			
	1	2	3	4
Нормальная сыворотка	0,316	0,328	0,214	0,252
Иммунная сыворотка	0,204	0,216	0,167	0,156

2. Изучалось продолжительность развития эмбриона (в днях) кроликов разных пород. Влияет ли породность на продолжительность развития эмбриона?

Породы	1	2	3	4	5	6	7	8	9	10
Альбиносы	30	36	31	30	34	32	34	32	33	32
Шиншилла	31	32	30	34	32	31	30	31	30	31
Голландские	30	29	30	31	30	30	31	30	31	31
Польские	30	31	29	30	29	30	29	31	29	30

3 Получены следующие данные о содержании хлорофилла (в мг/кв.дм) в листьях канатика в разное время суток. Влияет ли время суток на содержание хлорофилла?

Часы суток	1	2	3	4
15	3,06	2,88	2,83	2,41
18	3,20	2,97	2,50	3,03
21	1,82	1,73	1,33	2,25
24	1,67	1,26	1,52	1,36
6	2,76	1,26	1,46	1,32
9	2,78	2,7	2,49	1,66
12	2,41	3,22	1,9	2,00

4 Получены следующие данные о плодовитости самок мышей при облучении их рентгеновскими лучами. Влияет ли облучение на плодовитость мышей?

Группы	Число мышат от отдельных самок			
Доза 0 p	10	12	11	10
Доза 100 p	8	10	7	9

Доза 200р	7	9	6	4
-----------	---	---	---	---

5 Изучали процент гемоглобина в крови кур разных пород. Влияет ли породность на процент гемоглобина?

Породы	1	2	3	4	5	6	7
Итальянские	87	92	86	91	90	93	90
Куропатчатые	91	90	88	89	90	87	89
Минорки	85	82	85	86	89	84	85
Бентамы	82	82	85	83	82	83	84

6 Изучали живой вес ягнят при рождении (в кг), ношенных разное число дней:

Длительность Беременности	Живой вес ягнят									
	145	3,8	2,9	3,3	3,6	3,8	3,7	4,8	5,1	3,4
146	3,7	2,9	3,3	3,6	3,9	3,7	4,7	5,0	3,4	3,2
147	3,9	4,1	4,4	5,0	3,0	2,9	4,0	3,2	4,2	4,3
148	4,0	5,2	4,3	2,9	4,1	3,9	3,2	3,9	4,1	4,0
149	4,0	5,3	4,2	3,0	4,0	3,9	4,2	3,3	4,0	4,1
150	4,1	4,3	5,4	3,1	4,0	4,0	4,3	3,9	4,0	4,1
151	4,3	4,2	5,5	4,2	4,1	4,1	4,4	3,5	4,1	3,6
152	4,3	3,6	4,4	5,5	4,0	4,1	4,5	4,1	4,2	4,3
153	4,4	4,7	3,9	4,6	5,7	4,3	4,8	4,9	4,7	4,7

Примените метод дисперсионного анализа для выяснения влияния длительности плодоношения на живой вес ягнят.

7 .Используя метод дисперсионного анализа, определить достоверность влияния дозы микроэлемента на величину поглощения кислорода крысами.

Доза (А)	Ср. величина потребл. кислорода (в мл)		
А1	176	179	169
А2	162	167	168
А3	157	154	153

8 Методом дисперсионного анализа определить влияние форм клинического течения ревматизма на содержание гепарина в крови больных с ревматическим пороком сердца.

А-фаза течения болезни	Содержание гепарина.					
А1-острая	1	3	3	1	2	2
А2-вялая	2	2	3	3	4	4
А3-неактивная	2	4	3	4	5	6

9 Определяли содержание фосфатазы в эпифизах костей конечностей, подвергавшихся и не подвергавшихся однократной вибрации. Методом

дисперсионного анализа выяснить, влияет ли вибрация на изменение активности фосфатазы в костях конечности.

Уровни фактора	Содержание фосфатазы.						
Интактная конечность	43	30	63	47	39	33	45
Контактная конечность	72	35	83	70	44	51	59

10 Больных острым инфарктом миокарда в первый месяц лечения наряду с общепринятым лечением назначили ежедневный приём аспирина в разных дозировках. При этом оценивали снижение относительного риска смерти через 30 дней от начала лечения острого инфаркта миокарда. Влияет ли на эффективность лечения острого инфаркта миокарда назначение различных доз аспирина?

Суточная дозировка аспирина, мг/сут	№ испытания					
	1	2	3	4	5	6
75	5	9	14	17	18	16
160	21	24	26	31	33	22
325	22	33	24	26	29	31
500	14	17	27	21	22	25
1500	15	21	24	28	26	20

11 У испытуемых было изучено потребление кислорода (в MET) при различной физической активности. Влияет ли уровень физической активности на потребление кислорода?

Ходьба(км/ч)	№ испытания				
	1	2	3	4	5
1.5	2.5	2.4	2.7	2.2	2.6
3	3.1	3.3	2.9	3.0	2.9
5	4.9	5.4	5.2	5.7	5.3
6.5	5.8	6	5.7	5.4	5.1

12 При обострениях хронической обструктивной болезни лёгких используют лекарственный препарат будесонид. В таблице представлены значения парциального напряжения углекислого газа крови в зависимости от длительности терапии. Влияет ли продолжительность лечения будесонидом на парциальное напряжение углекислого газа крови?

дни	№ испытания				
	1	2	3	4	5
2	44.2	43.9	44.1	44	43.8
4	43.7	43.1	43.5	43.9	43
7	41.6	42	41.5	41.9	41.2

10	40.1	40.7	40.4	40.9	41
-----------	------	------	------	------	----

13 В исследовании изучали изменение вязкости цельной крови больных стенокардией II и III функционального класса под влиянием ЭЛМ излучения КВЧ-диапазона на частоте молекулярного спектра излучения и поглощения атмосферного кислорода с различной продолжительностью периода облучения образца крови. Влияет ли продолжительность облучения на вязкость крови?

Продолжительность облучения крови (мин)	№ испытания					
	1	2	3	4	4	5
0-15	5.4	5	4.5	5.1	4.7	4.9
15-30	4.6	4.4	4.0	4.3	4.2	4.6
30-60	3.5	3.7	4	3.4	3	3.3
60-80	3	3.1	3.2	3.4	3.3	3.1

14 больных острым инфарктом миокарда в различные дни от начала заболевания определяли количество эритроцитов. В таблице представлены значения эритроцитов в различные сроки от начала острого инфаркта миокарда. Влияет ли продолжительность заболевания на содержание эритроцитов в крови.

Продолжительность заболевания, дни	№ испытания					
	1	2	3	4	5	6
1	4.2	4.1	4.8	4.5	4	4.5
7	5.2	5.3	5.0	4.9	5.1	4.8
21	4.2	4.4	4.7	4.9	4.6	4.1

15 Оцените эффективность влияния небиволола на максимальную скорость кровотока в плечевой артерии(в м/с) через 6 мес. лечения у пациентов с сердечной недостаточностью.

Доза небивал.	№ испытания					
	1	2	3	4	5	6
1.25	0.34	0.32	0.33	0.35	0.34	0.32
2.5	0.54	0.53	0.55	0.56	0.54	0.53
5	0.61	0.63	0.64	0.62	0.63	0.66

16 В таблице отображены показатели фракции выброса левого желудочка у больных с хронической недостаточностью кровообращения различных функциональных классов. Определите, влияет ли функциональный класс недостаточности кровообращения на сократительную способность левого желудочка.

Фактор А	№ испытаний				
	1	2	3	4	5
1	0.47	0.45	0.41	0.4	0.43

2	0.48	0.43	0.41	0.42	0.41
3	0.33	0.32	0.34	0.3	0.35
4	0.23	0.21	0.2	0.24	0.25

17 В таблице отображены показатели индекса массы миокарда левого желудочка (г/м^2) у больных с хронической недостаточностью кровообращения различных функциональных классов. Определите, влияет ли функциональный класс недостаточности кровообращения на массу левого желудочка?

Функциональный класс	№ испытания				
	1	2	3	4	5
1	140	141	142	145	141
2	138	139	142	140	143
3	190	187	192	189	191
4	250	252	255	254	247

18 Проверьте эффективность влияния оликарда на количество приступов стенокардии в сутки после курсового лечения пациентов с ранней постинфарктной стенокардией.

Доза оликарда (мг/сут)	№ испытания					
	1	2	3	4	5	6
40	2	1	3	5	2	1
60	3	4	2	1	5	5
80	1	1	2	1	3	1

19 Проверьте, влияет ли степень тяжести (X) хронической обструктивной болезни лёгких на объём форсированного выхода за 1 сек. (в % от должного).

X	№ испытания					
	1	2	3	4	5	6
Лёгкая	70	75	74	80	72	76
Средняя	61	56	62	60	53	52
Тяжёлая	45	49	50	45	47	42

20 Проверьте, влияет ли возраст на частоту распространённости изолированной систолической артериальной гипертензии в различных регионах России (в %).

Возраст	Регионы				
	1	2	3	4	5
50	24	23	21	25	23
60	47	45	43	42	46
70	66	60	65	65	65
80	73	70	72	71	73

21 При уровне значимости $\alpha = 0,05$ методом дисперсионного анализа проверить значимость влияния двух факторов: pH среды (фактор *A*) и концентрации (в процентах) лиганда (фактор *B*) на экстракцию комплекса металла с лигандом из водной в органическую фазу.

Уровни A	Уровни B		
	B ₁ = 1	B ₂ = 2	B ₃ = 3
3	63	42	71
6	69	73	82
9	91	68	60

22 При уровне значимости $\alpha = 0,05$ методом дисперсионного анализа проверить значимость влияния двух факторов: pH среды (фактор *A*) и концентрации (в процентах) лиганда (фактор *B*) на экстракцию комплекса металла с лигандом из водной в органическую фазу.

Уровни A	Уровни B		
	B ₁ = 1	B ₂ = 2	B ₃ = 3
3	63	42	71
6	69	73	82
9	91	68	60

23 При уровне значимости $\alpha = 0,05$ методом дисперсионного анализа проверить существенность влияния температуры (фактор *A*) и фермента (фактор *B*) на выход продукта биохимического синтеза.

Уровни A	Уровни B	
	B ₁	B ₂
A ₁	100	90
A ₂	86	87