

# Информационные технологии и управление в области безопасности жизнедеятельности

УДК 311+519.23/.25

## Двухфакторный дисперсионный анализ гибели людей при пожарах в регионах Южного федерального округа

## Two-factor analysis of variance of death of people in fires in the regions of the southern federal district

*И.А. Кайбичев,  
д-р физ–мат. наук, доцент,  
К.И. Калимуллина  
ФГБОУ ВО Уральский инсти-  
тут ГПС МЧС России*

*I.A. Kaibichev,  
Holder of an Advanced  
Doctorate in Physico-  
mathematical Sciences,  
K.I. Kalimullina  
The Ural Institute of State  
Firefighting Service of Ministry  
of Russian Federation for Civil  
Defense*

### Аннотация:

Выполнен дисперсионный анализ гибели людей при пожарах в регионах Южного федерального округа по данным 2001-2016 годов. С вероятностью 0,95 доказано, что средние числа погибших для регионов Южного федерального округа различны. Различие обусловлено фактором географического положения. Средние числа погибших при пожарах по Южному федеральному округу по годам также различны. Различие обусловлено фактором времени. Установленные факты могут быть полезны при математическом моделировании обстановки с гибелью людей при пожарах в регионах Южного федерального округа.

**Ключевые слова:** количество погибших при пожарах, дисперсионный анализ, Южный федеральный округ.

### Abstract:

The dispersion analysis of deaths in fires in the regions of the Southern Federal district was performed according to the data of 2001-2016. With a probability of 0.95 it is proved that the average number of deaths in fires for the regions of the Southern Federal district are different. The difference is due to geographical location. The average number of deaths in fires in the South Federal district also varies from year to year. The difference is due to the time factor. The established facts can be useful in mathematical modeling of the situation with the death of people in fires in the regions of the South Federal district.

**Key words:** number of death in fires, dispersion analysis, Southern federal district.

Дисперсионный анализ является одним из методов математической статистики [1,2]. Он направлен на поиск зависимостей в данных на основе исследования значимости различий в средних значениях. Суть метода состоит в изучении влияния одной или нескольких переменных (обычно называемых факторами) на исследуемую переменную. Ранее дисперсионный анализ был применен для изучения обстановки с гибелью людей при пожарах в федеральных округах РФ [3].

В роли исследуемой переменной в данном исследовании выступает число погибших при пожарах в регионах Южного федерального округа.

В качестве факторов рассмотрим 2 порядковые переменные: условный номер региона и номер года. Поскольку в Южном федеральном округе всего 6 регионов, то целесообразно будет нумеровать их целыми числами от 1 до 6. Это позволит перейти от шкалы наименований (названий регионов) к порядковой шкале. Отметим, что номер года фигурирует во многих экономических исследованиях и отчетах Росстата. Под фактором времени в экономике принято понимать явление изменения экономических показателей, причиной которых является время. В основе действия фактора времени лежит изменение комплекса социально-экономических и технико-экономических условий. В результате одинаковые затраты труда, прикладываемые в одном и том же месте, но в различные периоды времени, либо приносят различные результаты, либо достижение одинаковых результатов достигается применением различных затрат труда. В нашем исследовании данные по гибели людей взяты из статистических сборников ВНИИПО [4-17] для 2001-2018 годов. Поэтому фактор времени будет целочисленной переменной в интервале от 2001 до 2018.

Выполним корреляционный анализ. Он позволит установить наличие зависимостей между числом погибших при пожарах и рассматриваемыми факторами. Расчет проводили с помощью программы Microsoft Excel. Коэффициент линейной корреляции Пирсона R вычисляли с помощью функции Коррел (Табл. 1).

Наличие всего 18 наблюдений привело к необходимости перерасчета коэффициента линейной корреляции Пирсона на случай малого объема данных [18-20]:

$$R' = R \left[ 1 + \frac{1-R^2}{2(n-3)} \right] \quad (1)$$

Очень сильная связь ( $0,95 \leq |R'| < 1$ ) наблюдается в Краснодарском крае и Волгоградской области. Ситуация сильной связи ( $0,75 \leq |R'| < 0,95$ ) имеет место в Астраханской, Ростовской областях, Республике Калмыкия, Республике Адыгея.

**Таблица 1 Коэффициент линейной корреляции Пирсона между числом погибших при пожарах и фактором времени в Южном федеральном округе**

Номер	Регион	R	R'
1	Республика Калмыкия	-0,87	-0,87
2	Республика Адыгея	-0,76	-0,77
3	Астраханская область	-0,93	-0,94
4	Волгоградская область	-0,95	-0,95
5	Ростовская область	-0,93	-0,94
6	Краснодарский край	-0,98	-0,98

В целом можно сделать вывод о наличии связи между числом погибших при пожарах и фактором времени (номером года).

Исследуем наличие связи между числом погибших при пожарах и номером региона (Табл. 2).

**Таблица 2 Коэффициент линейной корреляции Пирсона между числом погибших при пожарах и номером региона**

Год	R	R'
2001	0,96	0,97
2002	0,94	0,96
2003	0,95	0,96
2004	0,95	0,96
2005	0,94	0,96
2006	0,95	0,97
2007	0,95	0,97
2008	0,96	0,97
2009	0,96	0,97
2010	0,96	0,97
2011	0,96	0,97
2012	0,96	0,98
2013	0,97	0,98
2014	0,96	0,98
2015	0,97	0,98
2016	0,98	0,99
2017	0,97	0,98
2018	0,97	0,98

На основании величины R' для 2001-2018 годов можно сделать вывод о наличии очень сильной связи между числом погибших и номером региона.

Связь между числом погибших при пожарах и факторами времени и номером региона установлена. Возникает вопрос о достоверности полученных результатов. Для этого выполним проверку уровня значимости коэффициента корреляции. Формулируем две гипотезы. Гипотеза H0 состоит в том, что коэффициент корреляции между рассматриваемыми

величина равен 0 (связи нет). В альтернативной гипотезе H1 коэффициент корреляции отличен от нуля (связь есть).

Процедура проверки гипотез содержит расчет статистики Фишера

$$u = \frac{1}{2} \ln \frac{1+R'}{1-R'} \quad (2)$$

В Microsoft Excel мы использовали для этого функцию Фишер (Табл. 3). На следующем этапе задавали уровень значимости  $\alpha$  и вычисляли критические значения

$$u_{\alpha}(n) = z_{1-\frac{\alpha}{2}} \left( \frac{1}{\sqrt{n-3}} \right) \quad (3)$$

Здесь  $z_{1-\alpha/2}$  - квантили нормированного распределения, принимают значения 1,96 при уровне значимости  $\alpha = 0,05$  и 2,576 при  $\alpha = 0,01$ . В случае проверки гипотезы об уровне значимости коэффициента корреляции числа погибших при пожарах с фактором времени получили  $u_{0,05}(18) = 0,51$ ,  $u_{0,01}(18) = 0,67$ . В случае  $|u| \leq u_{\alpha}(n)$  принимается гипотеза H0. В ситуации случае  $|u| > u_{\alpha}(n)$  верна гипотеза H1.

**Таблица 3 Проверка гипотезы об уровне значимости коэффициента линейной корреляции Пирсона между числом погибших при пожарах и фактором времени**

Номер	Регион	u	Гипотеза ( $\alpha = 0,05$ )	Гипотеза ( $\alpha = 0,01$ )
1	Республика Калмыкия	-1,35	H1	H1
2	Республика Адыгея	-1,03	H1	H1
3	Астраханская область	-1,70	H1	H1
4	Волгоградская область	-1,85	H1	H1
5	Ростовская область	-1,71	H1	H1
6	Краснодарский край	-2,43	H1	H1

В итоге установлено, что при уровне значимости  $\alpha = 0,01$  (с вероятностью 0,99) коэффициент корреляции числа погибших при пожарах с фактором времени отличен от нуля.

При проверке гипотезы об уровне значимости коэффициента корреляции числа погибших при пожарах с номером региона рассчитанные значения статистики Фишера (Табл. 4) сравнивали с критическими значениями  $u_{0,05}(6) = 1,13$  и  $u_{0,01}(6) = 1,49$ .

**Таблица 4 Проверка гипотезы об уровне значимости коэффициента линейной корреляции Пирсона между числом погибших при пожарах и номером региона**

Год	u	Гипотеза ( $\alpha = 0,05$ )	Гипотеза ( $\alpha = 0,01$ )
2001	2,18	H1	H1
2002	1,97	H1	H1
2003	1,98	H1	H1
2004	1,99	H1	H1
2005	1,93	H1	H1
2006	2,03	H1	H1
2007	2,06	H1	H1
2008	2,11	H1	H1
2009	2,08	H1	H1
2010	2,18	H1	H1
2011	2,18	H1	H1
2012	2,20	H1	H1
2013	2,30	H1	H1
2014	2,20	H1	H1
2015	2,35	H1	H1
2016	2,45	H1	H1
2017	2,31	H1	H1
2018	2,21	H1	H1

Следовательно, при уровне значимости  $\alpha = 0,01$  (с вероятностью 0,99) коэффициент корреляции числа погибших при пожарах с номером региона отличен от нуля.

Проведенный корреляционный анализ показал обоснованность выбора факторов времени и номера региона. Число погибших при пожарах зависит от этих величин с вероятностью 0,99.

Перейдем к процедуре дисперсионного анализа. Фактор времени обозначим буквой А. Фактор номера региона обозначим буквой В.

Рассчитываем средние значения для регионов Южного федерального округа (Табл. 5, столбец Bj) и для годов по Южному федеральному округу (Табл. 5, строка Ai), общее среднее (в нашем случае 180).

**Таблица 5 Данные по числу погибших при пожарах в Южном федеральном округе**

№	Регион	2001	2002	2003	2004	2005	2006	2007
1	Республика Калмыкия	19	27	24	18	17	12	14
2	Республика Адыгея	48	23	41	25	24	43	36
3	Астраханская область	85	102	99	97	96	94	92
4	Волгоградская область	319	377	381	369	363	323	337
5	Ростовская область	417	414	375	387	450	432	405
6	Краснодарский край	468	459	468	449	437	416	427
Ai		226	234	231	224	231	220	219

№	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Bj
1	13	12	11	11	10	9	9	11	4	10	7	13
2	36	31	29	25	26	18	18	21	15	15	11	27
3	92	78	79	78	77	60	59	59	55	48	48	78
4	324	287	260	252	249	173	201	179	159	137	152	269
5	395	353	309	300	284	264	261	235	217	205	210	329
6	421	372	344	334	329	295	289	285	278	236	226	363
Ai	214	189	172	167	163	137	140	132	121	109	109	180

Выдвигаем гипотезы  $H_0(A)$  средние значения для регионов равны,  $H_1(A)$  средние значения для регионов различны;  $H_0(B)$  средние значения в Южном федеральном округе для годов равны,  $H_1(B)$  средние значения в Южном федеральном округе для годов различны. Из сформулированных гипотез предстоит выбрать достоверные.

В дисперсионном анализе рассчитывают общую сумму квадрата отклонений значений от общего среднего ( $SS$ ). Затем разбивают её на составляющие, связанные с влиянием факторов  $A$  ( $SS_A$ ) и  $B$  ( $SS_B$ ), а также ошибки ( $SS_C$ ). Последнюю компоненту нельзя объяснить влиянием факторов  $A$  и  $B$ . Считают, что она обусловлена влиянием посторонних случайных причин, которые невозможно учесть.

С помощью инструмента Анализ данных Microsoft Excel выполним расчет. Из всех возможных вариантов анализа выберем Двухфакторный дисперсионный анализ без повторений. Полученный результат состоит из двух частей. В первой части (Рис. 1) приведен расчет средних значений и дисперсий по столбцам и строкам Таблицы 5. Некоторое отличие от средних значений, полученных в Таблице 5, связано с округлением.

	A	B	C	D	E
1	Двухфакторный дисперсионный анализ без повторений				
2					
3	<b>ИТОГИ</b>	<b>Счет</b>	<b>Сумма</b>	<b>Среднее</b>	<b>Дисперсия</b>
4	Строка 1	18	238	13,22222222	33,830065
5	Строка 2	18	485	26,94444444	108,87908
6	Строка 3	18	1398	77,66666667	339,64706
7	Строка 4	18	4842	269	7197,4118
8	Строка 5	18	5913	328,5	7025,5588
9	Строка 6	18	6533	362,9444444	6773,7026
10					
11	Столбец 1	6	1356	226	39621,6
12	Столбец 2	6	1402	233,6666667	41653,467
13	Столбец 3	6	1388	231,3333333	39155,467
14	Столбец 4	6	1345	224,1666667	39276,967
15	Столбец 5	6	1387	231,1666667	42938,167
16	Столбец 6	6	1320	220	36887,6
17	Столбец 7	6	1311	218,5	36685,1
18	Столбец 8	6	1281	213,5	34935,5
19	Столбец 9	6	1133	188,8333333	27720,567
20	Столбец 10	6	1032	172	22223,2
21	Столбец 11	6	1000	166,6666667	21044,667
22	Столбец 12	6	975	162,5	19833,1
23	Столбец 13	6	819	136,5	15772,3
24	Столбец 14	6	837	139,5	15833,5
25	Столбец 15	6	790	131,6666667	13703,467
26	Столбец 16	6	728	121,3333333	12917,867
27	Столбец 17	6	651	108,5	9697,1
28	Столбец 18	6	654	109	9893,6

Рис. 1. Расчет средних значений и дисперсий

Вторая часть (Рис. 2) содержит результаты дисперсионного анализа.

31 Дисперсионный анализ							
32	Источник вариации	SS	df	MS	F	P-Значение	F критическое
33	Строки	2252778,6	5	450555,7204	261,97328	4,52371E-50	2,321812253
34	Столбцы	218955,935	17	12879,76089	7,4888701	7,29396E-11	1,744298846
35	Погрешность	146187,565	85	1719,853704			
36							
37	Итого	2617922,1	107				

  

29 Дисперсионный анализ							
30	Источник вариации	SS	df	MS	F	P-Значение	F критическое
31	Строки	2196486,08	5	439297,217	315,2049	7,55629E-49	2,336575649
32	Столбцы	150972,667	15	10064,8444	7,22173546	1,7246E-09	1,80182463
33	Погрешность	104526,583	75	1393,68778			
34							
35	Итого	2451985,33	95				

Рис. 2. Результаты дисперсионного анализа

Общая сумма квадратов отклонений значений от общего среднего

$$SS = SS_A + SS_B + SS_e \quad (4)$$

В нашем случае в выражение (1) войдут величины  $SS = 2617922,1$ ,  $SS_A = 218955,94$ ,  $SS_B = 3352778,6$ ,  $SS_e = 146187,57$ .

Число значений а фактора А равно 18, количество значений b для фактора В равно 6. Число степеней свободы для фактора А  $df_A = a-1 = 17$ , для фактора В  $df_B = b-1 = 5$ , ошибки  $df = (a-1)*(b-1) = 85$ .

Определим дисперсии

$$MS_A = \frac{Q_A}{df_A} = 12879,76, MS_B = \frac{Q_B}{df_B} = 450555,72, MS_e = \frac{Q_e}{df_e} = 1719,85. \quad (5)$$

Вычислим F – статистику

$$F_A = \frac{MS_A}{MS_e} = 7,49, F_B = \frac{MS_B}{MS_e} = 261,97. \quad (6)$$

Задаем уровень значимости  $\alpha = 0.05$ . Критические значения F – статистики равны

$$F_{\alpha}(df_A, df_e) = 1,74, F_{\alpha}(df_B, df_e) = 2,32 \quad (7)$$

В рассмотренном случае выполняются неравенства

$$F_A > F_{\alpha}(df_A, df_e), F_B > F_{\alpha}(df_B, df_e) \quad (8)$$

Поэтому с вероятностью 0,95 нужно принять гипотезы  $H_1(A)$ ,  $H_1(B)$ . Это означает, что средние значения погибших при пожарах для регионов Южного федерального округа различны. Средние значения чисел погибших при пожарах по Южному федеральному округу для годов также различны.

Коэффициент детерминации для фактора В

$$R_B^2 = \frac{SS_B}{SS} = 0,8605 = 86,05\% \quad (9)$$

свидетельствует, что фактор номера региона объясняет 89,58 % различий в значениях числа погибших при пожарах от общего среднего.

Коэффициент детерминации для фактора А

$$R_A^2 = \frac{SS_A}{SS} = 0,0836 = 8,36\% \quad (10)$$

показывает, что фактор времени объясняет 8,36 % различий.

Коэффициент детерминации ошибки

$$R_e^2 = \frac{SS_e}{SS} = 0,0558 = 5,58\% \quad (8)$$

дает основания для утверждения, что необъясненной остается 5,58 % различий значений чисел погибших при пожарах от общего среднего.

С вероятностью 0,95 установлено, что средние числа погибших при пожарах для регионов Южного федерального округа различны. Различие вызвано влиянием фактора условного номера региона. С вероятностью 0,95 справедлива гипотеза о различии средних значений числа погибших при пожарах в Южном федеральном округе для годов. Различие обусловлено фактором времени. Основной вклад в различие чисел пожаров от общего среднего дает фактор условного номера региона.

Полученные результаты могут оказаться полезными при разработке математических моделей, объясняющих число погибших при пожарах в Южном федеральном округе.

## Литература

1. Дисперсионный анализ // ru.wikipedia.org. Дисперсионный анализ – Википедия.
2. Шеффе Г. Дисперсионный анализ. – М.: Наука, 1980. – 512 с.
3. Кабичев И.А., Кайбичева Е.И. Дисперсионный анализ гибели людей в федеральных округах Российской Федерации // Техносферная безопасность, 2019, № 1(22), с. 28-31.
4. Пожары и пожарная безопасность в 2005 году. Статистический сборник / под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2006. 139 с.
5. Пожары и пожарная безопасность в 2006 году: Статистический сборник / под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2007. 137 с.
6. Пожары и пожарная безопасность в 2007 году. Статистический сборник / под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2008. 137 с.
7. Пожары и пожарная безопасность в 2008 году. Статистический сборник / под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2009. 137 с.
8. Пожары и пожарная безопасность в 2009 году. Статистический сборник / под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2010. 135 с.
9. Пожары и пожарная безопасность в 2010 году. Статистический сборник / под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2011. 140 с.
10. Пожары и пожарная безопасность в 2011 году. Статистический сборник / под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2012. 137 с.
11. Пожары и пожарная безопасность в 2012 году. Статистический сборник / под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2013. 137 с.
12. Пожары и пожарная безопасность в 2013 году. Статистический сборник / под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2014. 137 с.
13. Пожары и пожарная безопасность в 2014 году. Статистический сборник / под общей редакцией А.В. Матюшина. – М.: ВНИИПО, 2015. 124 с.
14. Пожары и пожарная безопасность в 2015 году. Статистический сборник / под общей редакцией А.В. Матюшина. – М.: ВНИИПО, 2016. 124 с.
15. Пожары и пожарная безопасность в 2016 году. Статистический сборник / под общей редакцией Д.М. Гордиенко. – М.: ВНИИПО, 2017. 124 с.
16. Пожары и пожарная безопасность в 2017 году. Статистический сборник / под общей редакцией Д.М. Гордиенко. – М.: ВНИИПО, 2018. 125 с.
17. Пожары и пожарная безопасность в 2018 году. Статистический сборник / под общей редакцией Д.М. Гордиенко. – М.: ВНИИПО, 2019. 125 с.
18. Харченко М.А. Корреляционный анализ: учебное пособие для вузов. – Воронеж: Издательско-полиграфический центр Воронежского государственного университета, 2008. 31 с.
19. Рубан А.И. Методы анализа данных: учебное пособие. – Красноярск: ИПЦ КГТУ, 2004. 319 с.
20. Лакин Г.Ф. Биометрия. – М.: Высшая школа, 1990. 350 с.: