УДК 519.2

РЕШЕНИЕ ПРИКЛАДНЫХ ЗАДАЧ ЭКОНОМИКИ И ТЕОРИИ НАДЕЖНОСТИ МЕТОДОМ ДИСПЕРСИОННОГО АНАЛИЗА

Т. А. Певцова, кандидат физико-математических наук, доцент кафедры алгебры и геометрии факультета математики и информационных технологий ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет имени Н. П. Огарёва» (г. Саранск, Россия), pevtsovata@mail.ru
Е. А. Рябухина, кандидат педагогических наук, доцент кафедры систем автоматизированного проектирования факультета математики и информационных технологий ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет имени Н. П. Огарёва» (г. Саранск, Россия), ryabuhinaea@yandex.ru
О. А. Гущина, кандидат технических наук, доцент кафедры систем автоматизированного проектирования факультета математики и информационных технологий ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет имени Н. П. Огарёва» (г. Саранск, Россия), motuzoa@mail.ru

Рассматривается ряд стохастических методов, построенных на схеме независимых испытаний для решения широкого класса прикладных задач в экономике и теории надежности с привлечением математического аппарата матричного исчисления, комбинаторного и дисперсионного анализа.

Ключевые слова: реализация, модель, вероятность, критерий, надежность, дисперсионный анализ.

Интенсификация производства, широкое внедрение в экономику наукоемких технологий требуют применения математических моделей и средств их реализации с привлечением информационных технологий.

Цель исследования – представить эффективные методы прикладной статистики в решении задач экономики и теории надежности. Его актуальность обусловлена изменением требований к качеству управления производством. Современные методы принятия управленческих решений широко привлекают методы прикладной статистики, в частности дисперсионного анализа. Дисперсионный анализ является эффективным инструментом оценки влияния на объекты управления как отдельных факторов, так и их сочетаний.

Результаты исследования востребованы во всех отраслях материального производства и экономике. Поскольку становление производства базируется на развитии электроэнергетики, актуализируется задача обеспечения надежности функционирования энергетических объектов, решаемая средствами теории надежности.

В статье подробно рассмотрены аспекты применения методов дисперсионного анализа для реализации прикладных задач экономики и теории надежности, которые позволяют значительно повысить эффективность решения задач, приведенных в литературе [1–5].

Обратимся к серии прикладных задач на применение дисперсионного анализа.

Пример 1. В табл. 1 представлено время X проведения ремонта производственных мощностей по четырем разным технологиям (k = 4 – количество групп (градаций)). Требуется исследовать данный признак X методами дисперсионного анализа.

Таблица 1

Факторный признак X			
X_1	X_2	X_3	X_4
1,2	3,6	4	9,2
8	2,6	10	8
11,2	8	9,2	7

Признак X равномерно градуирован в четыре группы (r = 4), объемы которых одинаковы (N = 3), следовательно, объем всей выборки $n = 3 \times 4 = 12$. Составим расчетную таблицу (табл. 2).

Таблица 2

Показатели	***Факторный признак X***				***Суммы***
	X_1	X_2	X_3	X_4	
1	2	3	4	5	6
	1,2	3,6	4	9,2	
	8	2,6	10	8	**r = 4**
	11,2	8	9,2	7	
					n = 12
N	3	3	3	3	**12**

Окончание табл. 2

1	2	3	4	5	6
$\sum x_i$	20,4	14,2	23,2	24,2	**82**
$(\sum x_i)^2$	416,16	201,64	538,24	585,64	**1 741,68**
$(\sum x_i)^2$ / N	138,72	67,213	179,41	195,21	**580,56**
$\sum x_i^2$	190,88	83,72	200,64	197,64	**672,88**

1. Для вычисления дисперсий воспользуемся расчетными формулами:

$$D_O = \sum_{i=1}^{n} x_i^{\,2} - H, \text{ где } H = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}; \quad D_X = \frac{\sum_{i=1}^{n} (\sum x_i)^2}{N} - H; \; D_e = D_O - D_X.$$

Сначала определим величину H: $H = \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \dfrac{82^2}{12} = 560{,}33.$

Вычислим общую дисперсию D_O и составим расчетную таблицу (табл. 3).

Таблица 3

X_1	X_2	X_3	X_4	X_1^2	X_2^2	X_3^2	X_4^2
1,2	3,6	4	9,2	1,44	12,96	16	84,64
8	2,6	10	8	64	6,76	100	64
11,2	8	9,2	7	125,44	64	84,64	49
				190,88	83,72	200,64	197,64
$\sum$				**672,88**			

В последней строке табл. 3 находится величина $\sum_{i=1}^{n} x_i^{\,2} = 672{,}88.$

Определим общую дисперсию D_O: $D_O = \sum_{i=1}^{n} x_i^{\,2} - H = 672{,}88 - 560{,}33 = 112{,}55.$

Так как $\dfrac{\sum_{i=1}^{n} (\sum x_i)^2}{N} = 580{,}56$, то $D_X = \dfrac{\sum_{i=1}^{n} (\sum x_i)^2}{n} - H = 580{,}56 - 560{,}33 = 20{,}23.$

Затем выясним остаточную дисперсию $D_e = D_O - D_X = 112{,}55 - 20{,}23 = 92{,}32.$

2. Находим число степеней свободы для каждой дисперсии:

$k_o = n - 1 = 12 - 1 = 11;$

$k_X = k - 1 = 4 - 1 = 3;$

$k_e = n - k = 12 - 4 = 8$.

3. Вычисляем соответствующие исправленные дисперсии:

а) исправленная общая дисперсия $S^2_{\text{О}} = D_{\text{О}} / k_{\text{О}} = \frac{112{,}55}{11} \approx 10{,}23$;

б) факториальная исправленная межгрупповая дисперсия $S^2_X = \frac{D_X}{k_X} = \frac{20{,}23}{3} \approx$ 6,743;

в) факториальная исправленная внутригрупповая (остаточная) дисперсия

$S^2_{\text{е}} = \frac{D_e}{k_e} = \frac{92{,}32}{8} = 11{,}54$.

4. Для сравнения дисперсий вычисляем наблюдаемое значение критерия Фишера–Снедекора ($F_{\text{НАБЛ}}$) как отношение дисперсии $S^2_{\text{е}}$ к S^2_X:

$$F_{\text{НАБЛ}} = \frac{S^2_X}{S^2_e} = \frac{6{,}743}{11{,}54} \approx 0{,}584.$$

Критическое значение F-критерия ($F_{\text{КРИТ}}$) можно найти:

а) по таблице «Критические точки распределения Фишера–Снедекора» для уровня значимости $\alpha = 0{,}05$ и степенях свободы

$k_1 = k_X = 3$ и $k_2 = k_e = 8$: $F_{\text{КРИТ}} = F(\alpha; k_1; k_2) = F(0{,}05; 3; 8) = 4{,}07$;

б) воспользовавшись возможностями табличного процессора MS Excel, а именно стандартной процедурой FРАСПОБР:

$F_{\text{КРИТ}}$ = FРАСПОБР(α; k_1; k_2) = FРАСПОБР(0,05; 3; 8) = 4,066.

Теперь сравним наблюдаемое и критическое значения F-критерия. Так как неравенство $F_{\text{НАБЛ}} < F_{\text{КРИТ}}$ выполняется, то нулевая гипотеза о равенстве групповых средних принимается.

Вывод: согласно расчетам методом дисперсионного анализа с вероятностью 95 % можно предположить, что время проведения ремонта производственных мощностей по четырем разным технологиям существенно не различается.

Для подтверждения результатов расчетов с помощью дисперсионного анализа воспользуемся стандартной процедурой однофакторного анализа MS Excel: Сервис \ Анализ данных \ Однофакторный дисперсионный анализ.

В качестве «*Входного интервала*» укажем диапазон данных; флажок из пункта «*Метки в первой строке*» должен отсутствовать; в пункте «*Группирование»* в подпункте *«по столбцам»*» нужно установить указатель; в качестве «*Выходного интервала*» укажем любую свободную ячейку, которая будет располагаться левее и выше ячейки диапазона результатов. Мы получим таблицу с итогами (табл. 4).

Таблица 4

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
Столбец 1	3	20,4	6,80	26,08		
Столбец 2	3	14,2	4,73	8,25		
Столбец 3	3	23,2	7,73	10,61		
Столбец 4	3	24,2	8,07	1,21		
Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-значение*	*F критическое*
Между группами	**20,23**	3	6,742	0,584	0,642	4,066
Внутри групп	**92,32**	8	11,54			
Итого	**112,55**	11				

Совпадение результатов приведенных вычислений и дисперсионного анализа подтверждает правильность расчетов.

Для оценки достоверности результатов о влиянии фактора на признак можно воспользоваться критерием Снедекора, если комплекс данных является равномерным (вариант А), либо критерием Плохинского, если комплекс данных не является равномерным (вариант Б); затем оценим значимость разницы средних для каждой градации фактора.

А) Так как комплекс данных равномерный (объемы групп–градаций равны), проверяем достоверность вывода по Снедекору. Для этого вычисляем величину $h_X{}^2$:

$$h_X{}^2 = \frac{\left|S_X{}^2 - S_e{}^2\right|}{S_X{}^2 + (N-1)S_e{}^2},$$

где N – число вариантов в каждой градации.

С учетом значений дисперсий S_X^2 = 6,743; S_e^2 = 11,54 и N = 3 вычисляем

$$h_X{}^2 = \frac{\left|S_X{}^2 - S_e{}^2\right|}{S_X{}^2 + (N-1)S_e{}^2} = \frac{|6{,}743 - 11{,}54|}{6{,}743 + (3-1)\cdot 11{,}54} = \frac{4{,}797}{29{,}823} = 0{,}161.$$

Далее найдем значение $F_{НАБЛ}$ критерия достоверности:

$$F_{НАБЛ} = \frac{h_X{}^2(n-r)}{(1-h_X{}^2)(r-1)}.$$

Учитывая, что n = 12; r = 4; $h_X{}^2$ = 0,161, вычисляем:

$$F_{НАБЛ} = \frac{h_X{}^2(n-r)}{(1-h_X{}^2)(r-1)} = \frac{0{,}161\cdot(12-4)}{(1-0{,}161)(4-1)} = \frac{0{,}161\cdot 8}{0{,}939\cdot 3} = \frac{1{,}288}{2{,}517} = 0{,}512.$$

Сравниваем величины $F_{НАБЛ}$ и $F_{КРИТ.}$ Поскольку выполняется неравенство $F_{НАБЛ} < F_{КРИТ}$, то с вероятностью 95 % можно считать, что влияние фактора на признак недостоверно и вывод, сделанный ранее, подтверждается, т. е. существенной разницы между выборочными средними градаций нет.

Б) Оценим значимость разницы средних для каждой градации фактора в комплексе. Для этого по формуле Тьюки вычислим значение критерия $T_{НАБЛ}$:

$$T_{НАБЛ} = \frac{|x_1{}^* - x_2{}^*|}{\sqrt{S_e^2 / k_e}},$$

где x_1* и x_2* – выборочные средние в сравниваемых градациях.

Так как x_1* = 6,80; x_2* = 4,733; S_e^2 = 11,54; k_e = 8, получаем

$$T_{НАБЛ} = \frac{|x_1{}^* - x_2{}^*|}{\sqrt{S_e^2 / k_e}} = \frac{|6{,}80 - 4{,}733|}{\sqrt{11{,}54/8}} = \frac{2{,}067}{\sqrt{1{,}4425}} = \frac{2{,}067}{1{,}201} = 1{,}721.$$

Критическое значение t-критерия можно найти по таблице «Критические точки распределения Стьюдента» $T_{КРИТ}$ = Т(α; k), а также с помощью стандартной процедуры FРАСПОБР табличного процессора MS Excel: $T_{КРИТ}$ = СТЬЮДРАСПОБР(α; k) для заданного уровня значимости α = 0,05 и числа степеней свободы $k = k_e$ = 8:

$T_{КРИТ}$ = Т(α; k) = Т(0,05; 8) = 2,31, или

$T_{КРИТ}$ = СТЬЮДРАСПОБР(α; k) = СТЬЮДРАСПОБР(0,05; 8) = 2,306.

Проверка. Если выполняется неравенство $T_{НАБЛ} > T_{КРИТ}$, то различия средних выборочных в двух рассматриваемых градациях признака считаются достоверными, и наоборот.

Так как $Т_{НАБЛ}$ = 1,721 и $T_{КРИТ}$ = 2,306, то неравенство $T_{НАБЛ} > T_{КРИТ}$ не выполняется и подтверждается сделанный ранее вывод об отсутствии существенной разницы между временем проведения ремонта производственных мощностей по четырем разным технологиям.

Различия выборочных средних проверяются только для попарно взятых градаций, притом критерий Стьюдента используется однократно.

Рассмотрим пример на технологию применения методов дисперсионного анализа для *неравномерного* нормально распределенного комплекса данных.

Пример 2. В табл. 5 указано время X проведения ремонта производственных мощностей по четырем разным технологиям.

Требуется исследовать данный признак X, представляющий неравномерный комплекс данных, который распределен нормально, методами дисперсионного анализа.

Таблица 5

Факторный признак X			
X_1	X_2	X_3	X_4
8	8,2; 9	11	7,5
8,4	10; 10	13	8,5
9	9,2; 10	12	
8,6			

Для изучения признака X, неравномерно градуированного в четыре группы (r = 4), объемы которых разные (N_1 = 4, N_2 = 6, N_3 = 3, N_4 = 2), удобно воспользоваться компьютерными технологиями из стандартного приложения MS WINDOWS – табличным процессором MS Excel.

Составим расчетную таблицу (табл. 6): N_j находим с помощью стандартной функции СЧЕТ; $\sum x_i$ и $\sum x_i^2$ – с помощью функции СУММ; $(\sum x_i)^2$ /

N_i – с помощью функции СРЗНАЧ.

Таблица 6

	Факторный признак X				**Суммы**
	X_1	X_2	X_3	X_4	
1	**2**	**3**	**4**	**5**	**6**
	8	8,2; 9	11	7,5	
	8,4	10; 10	13	8,5	***r* = 4**
	9	9,2; 10	12		
	8,6				
					n = 15
N_i	**4**	**6**	**3**	**2**	**15**
$\sum x_i$	34	56,4	36	16	**142,4**
$(\sum x_i)^2$	1 156	3 180,96	1296	256	**5 888,96**
$(\sum x_i)^2 / N_i$	289	530,16	432	128	**1 379,16**
$\sum x_i^2$	208,52	532,88	434	128,5	**1 384,90**

1. Для вычисления дисперсий воспользуемся расчетными формулами:

$$D_O = \sum_{i=1}^{n} x_i^{\,2} - H \text{, где } H = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}; \quad D_X = \sum \frac{\sum x_i^{\,2}}{N_i} - H \text{, } D_e = D_O - D_X;$$

$$H = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = \frac{142{,}4^2}{15} = \frac{20277{,}76}{15} = 1\,351{,}861;$$

$$D_O = \sum_{i=1}^{n} x_i^{\,2} - H = 1\,384{,}90 - 1\,351{,}861 = 33{,}05;$$

$D_X = 1\,379{,}16 - 1\,351{,}861 = 27{,}31;$

$D_e = D_O - D_X = 33{,}05 - 27{,}31 = 5{,}74.$

2. Находим число степеней свободы:

$k_o = n - 1 = 15 - 1 = 14;$

$k_X = k - 1 = 4 - 1 = 3;$

$k_e = n - k = 15 - 4 = 11.$

Вычисляем исправленные дисперсии:

$$S^2_X = \frac{D_X}{k_X} = \frac{27{,}31}{3} \approx \underline{9{,}103} \text{ и } S^2_e = \frac{D_e}{k_e} = \frac{5{,}74}{11} = 0{,}52.$$

3. Для сравнения дисперсий вычисляем наблюдаемое значение *F*-

критерия Фишера – Снедекора, а именно отношение большей дисперсии к меньшей. Так как $S^2_X > S^2_e$, то $F_{НАБЛ} = \frac{S^2_X}{S^2_e} = \frac{9{,}103}{0{,}52} \approx 17{,}51$.

По таблице «Критические точки распределения F Фишера – Снедекора» находим $F_{КРИТ}$ при уровне значимости $\alpha = 0{,}05$ и степенях свободы $k_1 = k_X = 3$ и $k_2 = k_e = 11$:

$F_{КРИТ} = F(\alpha; k_1; k_2) = F(0{,}05; 3; 11) = 3{,}59$;

или $F_{КРИТ}$ = FРАСПОБР(α; k_1; k_2) = FРАСПОБР(0,05; 3; 11) = 3,587.

Поскольку выполняется неравенство $F_{НАБЛ} > F_{КРИТ}$, то нулевая гипотеза о равенстве групповых средних отвергается.

Вывод. Согласно дисперсионному анализу приведенных данных с вероятностью 95 % можно предположить, что разница в длительности времени ремонта производственных мощностей по четырем разным технологиям является существенной. Другими словами, влияние фактора технологии на продолжительность ремонта доказано с помощью критерия.

Рекомендуется при проверке результатов дисперсионного анализа на достоверность не ограничиваться применением только одного критерия, а подтверждать выводы с помощью всевозможных других критериев:

1) по Плохинскому можно определить степень влияния факторов на не равномерно градуированный признак;

2) по Снедекору можно оценить достоверность полученных результатов;

3) t-критерий Стьюдента дает возможность оценить значимость разницы средних для каждой градации фактора.

Так как признак в *примере* 2 неравномерно градуированный, определяем степень влияния факторов по критерию Плохинского. Для этого нужно вычислить величину

$$h_X^2 = \frac{D_O}{D_e}, \quad h_X^2 = \frac{33{,}05}{5{,}74} = 5{,}76.$$

С помощью h_X^2 найдем наблюдаемое значение F-критерия достоверности: $F_{НАБЛ} = \left| \frac{h_X^2(n-r)}{(1-h_X^2)(r-1)} \right|$,

при значениях n = 15, r = 4 вычисляем:

$$F_{НАБЛ} = \left| \frac{h_X^2(n-r)}{(1-h_X^2)(r-1)} \right| = \left| \frac{5{,}76 \cdot (15-4)}{(1-5{,}76)(4-1)} \right| = \left| \frac{5{,}76 \cdot 11}{(-4{,}76) \cdot 3} \right| = \frac{63{,}36}{14{,}28} = 4{,}44.$$

Выше было найдено критическое значение $F_{КРИТ}$ = 3,59. Поскольку выполняется неравенство $F_{НАБЛ} > F_{КРИТ}$, то влияние фактора на признак по критерию Плохинского считается достоверным.

Проведенное исследование демонстрирует практическое применение дисперсионного анализа для решения задач оценивания надежности функционирования экономических объектов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Копыл В. И.* Логистика управления запасами с помощью Excel: производственно-практическое издание [Электронный ресурс] / В. И. Копыл. – Минск: Харвест, 2007. – 64 с. – Режим доступа: http://kutuzov.ua/product/70715/.

2. *Папков Б. В.* Надежность и эффективность электроснабжения / Б. В. Папков. – Н. Новгород: НГТУ, 1996. – 212 с.

3. *Папков Б. В.* Надежность электроснабжения: комплекс учеб.-метод. материалов [Электронный ресурс] / Б. В. Папков. – Н. Новгород: НТБ НГТУ, 2007. – 212 с. – Режим доступа: http://www.studmed.ru/papkov-bv-nadezhnost-elektrosnabzheniya-kompleks-uchebno-metodicheskih-aterialov_a83896f0bae.html.

4. *Певцова Т. А.* Вычисление индекса сезонности [Электронный ресурс] / Т. А. Певцова, Е. А. Рябухина, О. А. Гущина // Вестник Мордовского университета. – 2015. – Т. 25, № 4. – С. 18–36. – Режим доступа: http://cyberleninka.ru/article/n/vychislenie-indeksa-sezonnosti.

5. *Певцова Т. А.* Сетевые модели в теории надежности [Электронный

ресурс] // Studium. – 2016. – № 2 (39). – Режим доступа: http://saransk.ruc.su/upload/medialibrary/845/pevcova_ryabuhina_gushina01.pdf . (Дата обращения: 2.03.2018)