

РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

Факультет физико-математических и естественных наук

Кафедра прикладной информатики и теории вероятностей

Практикум по математической статистике

Лабораторная работа №3

Тема: «Дисперсионный анализ»

Вариант 10

Выполнил

Студент: Феокистов Владислав

Группа: НПМбд-01-196

№ с/б: 1032192939

Преподаватель: Матюшенко Сергей Иванович

МОСКВА

2022 г.

Цель работы: приобрести практические навыки применения дисперсионного анализа для решения конкретных задач с использованием статистического пакета SPSS.

Ход работы:

1. Повторил теоретические основы дисперсионного анализа, используя материалы учебного пособия.
2. Разобрал примеры использования SPSS для реализации дисперсионного анализа.
3. В приведенной ниже таблице даны значения урожайности картофеля (тыс. тонн с гектара) в зависимости от сорта картофеля (фактор A) и типа примененного удобрения (фактор B). С помощью двухфакторного дисперсионного анализа без повторных измерений необходимо выяснить:
 - a. значимы ли различия в средней урожайности различных сортов картофеля независимо от типа удобрения;
 - b. значимо ли влияние типа применяемого удобрения на урожайность независимо от сорта.

	B ₁	B ₂	B ₃	B ₄
A ₁	2.609	3.055	5.013	4.53
A ₂	4.191	4.452	5.683	4.647
A ₃	6.067	5.446	3.665	5.621

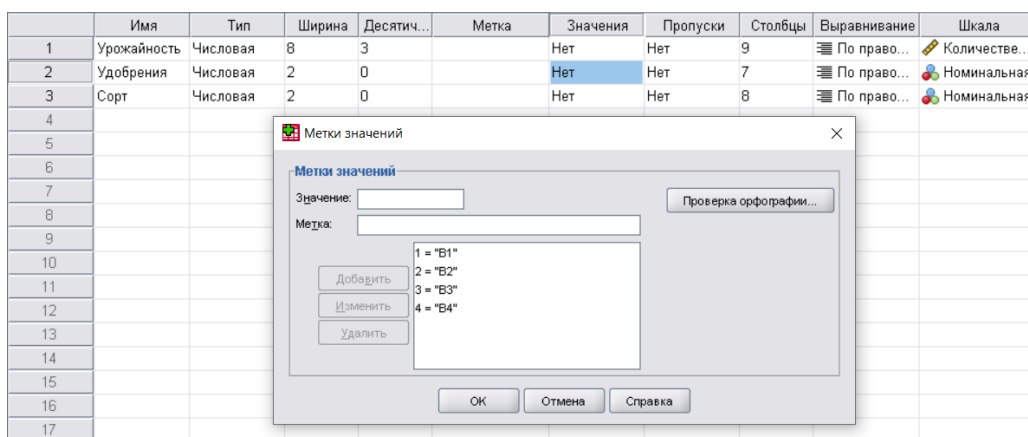
Решение. Согласно условию задачи исследуется влияние на урожайность (зависимую переменную) двух факторов – типа удобрений и сорта пшеницы.

Выдвинем две нулевые гипотезы:

$H_{0,A}$: различия в средней урожайности картофеля, вызванные влиянием типа удобрения (фактора A), выражены не более, чем различия, обусловленные случайными причинами.

$H_{0,B}$: различия в средней урожайности картофеля, вызванные влиянием сорта

Для начала запустим программу SPSS и в окне «Переменные» опишем все исходные данные (для столбцов с номинальной шкалой можно добавить значения для соответствующих уровней).

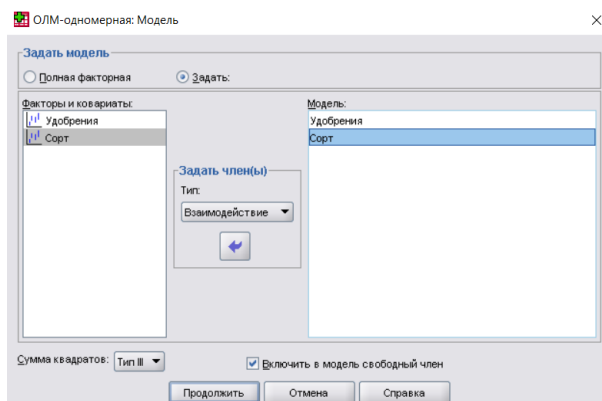
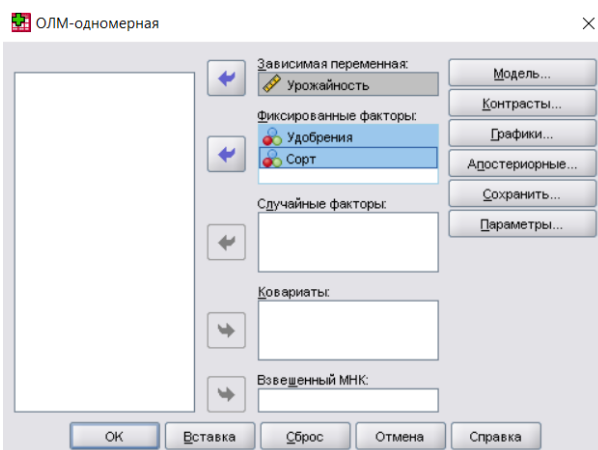


	Имя	Тип	Ширина	Десятич...	Метка	Значения	Пропуски	Столбцы	Выравнивание	Шкала
1	Урожайность	Числовая	8	3		Нет	Нет	8	По право...	Количество...
2	Удобрения	Числовая	2	0		{1, B1}...	Нет	8	По право...	Номинальная
3	Сорт	Числовая	2	0		{1, A1}...	Нет	8	По право...	Номинальная

В окне «Данные» введем значения переменных согласно условию задачи, используя вышеприведенную таблицу. Первый столбец исходных данных отведем для значений урожайности картофеля, второй – тип удобрения, третий – для сорта картофеля.

	Урожайность	Удобрения	Сорт
1	2,609	1	1
2	4,191	1	2
3	6,067	1	3
4	3,055	2	1
5	4,452	2	2
6	5,446	2	3
7	5,013	3	1
8	5,683	3	2
9	3,665	3	3
10	4,530	4	1
11	4,647	4	2
12	5,621	4	3

Теперь можно провести двухфакторный дисперсионный анализ без повторных изменений.



ОЛМ-одномерная: Графики профилей

Факторы:
Удобрения
Сорт

Горизонтальная ось:

Отдельные линии:

Отдельные графики:

Графики: Добавить Изменить Удалить

Удобрения*Сорт

Продолжить Отмена Справка

ОЛМ-одномерная: Параметры

Оцененные маргинальные средние

Факторы и их взаимодействия:
(OVERALL)
Удобрения
Сорт

Вывести средние для:
(OVERALL)
Удобрения
Сорт

☐ Сравнить главные эффекты

Корректировка доверительных интервалов:
LSD(нет)

Вывести среднее значение для...

☒ Описательные статистики
☒ Оценки размера эффекта
☐ Наблюдённая мощность
☐ Оценки параметров
☐ Матрица коэффициентов контрастов

☐ Критерии однородности
☐ График разброса по уровням
☐ Графики остатков
☐ Критерий отсутствия согласия
☐ Общая функция, допускающая оценку

Уровень значимости: 0,05 Доверительные интервалы: 95,0%

Продолжить Отмена Справка

ОЛМ-одномерная: Апостериорные множественные сравнения для наблюдаемых средних

Фактор(ы):
Удобрения
Сорт

Апостериорные критерии для:
Удобрения
Сорт

При равенстве дисперсий

☐ НЗР ☐ С-Н-К ☐ Уоллер-Дункан
☐ Бонферрони ☐ Тьюки ☐ Тьюки
☐ Шидак ☐ b Тьюки ☐ Даннетт
☒ Шеффе ☐ Дункан ☐ GT2 Гохберга
☐ F P-3-G-Y ☐ Габриэль
☐ Q P-3-G-Y

Отношение ошибок тип II: 100

Контрольная категория: Последний

Критерий:
☒ 2-сторонний ☐ < эталонного значения ☐ > эталонного значения

Равенство дисперсий не предполагается

☐ T2 Тамхойна ☐ T3 Даннетт ☐ Геймс-Хоуэлл ☐ С Даннетта

Продолжить Отмена Справка

После всех предварительных настроек «ОЛМ-одномерная», как показано на рисунках выше, нажимаем на кнопку «ОК» и получаем результаты расчета в виде таблиц и графиков.

Вначале выводится сводная таблица «Межгрупповые факторы», в которой приведены общие сведения об изучаемых факторах, присвоенных метках и о количестве наблюдений (N) по каждому фактору.

Одномерный дисперсионный анализ

Межгрупповые факторы

		Метка значения	М
Удобрения	1	B1	3
	2	B2	3
	3	B3	3
	4	B4	3
Сорт	1	A1	4
	2	A2	4
	3	A3	4

В таблице «Описательные статистики» содержатся средние значения и стандартные отклонения всех выборок, а также итоговые значения по всем данным.

Описательные статистики

Зависимая переменная: Урожайность				
Удобрения	Сорт	Для среднего	Стд. Отклонение	М
B1	A1	2,60900	.	1
	A2	4,19100	.	1
	A3	6,06700	.	1
	Всего	4,28900	1,731082	3
B2	A1	3,05500	.	1
	A2	4,45200	.	1
	A3	5,44600	.	1
	Всего	4,31767	1,201147	3
B3	A1	5,01300	.	1
	A2	5,68300	.	1
	A3	3,66500	.	1
	Всего	4,78700	1,027807	3
B4	A1	4,53000	.	1
	A2	4,64700	.	1
	A3	5,62100	.	1
	Всего	4,93267	,598978	3
Всего	A1	3,80175	1,151486	4
	A2	4,74325	,653759	4
	A3	5,19975	1,056041	4
	Всего	4,58158	1,073281	12

Таблица «Оценки эффектов межгрупповых факторов» содержит результаты проверки основных гипотез двухфакторного дисперсионного анализа.

В данном случае имеем следующее:

1. Переменная «Удобрение» оказывает статистически **значимое** влияние на распределение зависимой переменной «Урожайность», поскольку $F = 0,252$ при $\text{Знач.} = 0,858$ (средние значения урожайности по типам удобрений составили: 4,289; 4,318; 4,787 и 4,933).
2. Переменная «Сорт» также оказывает статистически **достоверное** влияние на распределение зависимой переменной «Урожайность», поскольку $F = 1,596$, а $\text{Знач.} = 0,278$ (средние значения урожайности по сортам: 3,802; 4,743 и 5,200).

Оценка эффектов межгрупповых факторов

Зависимая переменная: Урожайность

Источник	Сумма квадратов типа III	ст. св.	Средний квадрат	Щ	Знч.	Частная Эта в Квадрате
Скорректированная модель	5,028 ^a	5	1,006	,789	,593	,397
Свободный член	251,891	1	251,891	197,730	,000	,971
Удобрения	,962	3	,321	,252	,858	,112
Сорт	4,066	2	2,033	1,596	,278	,347
Ошибка	7,643	6	1,274			
Всего	264,562	12				
Скорректированный итог	12,671	11				

a. R квадрат = ,397 (Скорректированный R квадрат = -,106)

В следующих трех таблицах выводятся описательные статистики для совокупности всех данных и для каждого из исследуемых факторов.

Оцененные маргинальные средние

1. Среднее средних

Зависимая переменная: Урожайность

Для среднего	Стд. Ошибка	95% доверительный интервал	
		Нижняя граница	Верхняя граница
4,582	,326	3,784	5,379

2. Удобрения

Зависимая переменная: Урожайность

Удобрения	Для среднего	Стд. Ошибка	95% доверительный интервал	
			Нижняя граница	Верхняя граница
B1	4,289	,652	2,694	5,884
B2	4,318	,652	2,723	5,912
B3	4,787	,652	3,192	6,382
B4	4,933	,652	3,338	6,527

3. Сорт

Зависимая переменная: Урожайность

Сорт	Для среднего	Стд. Ошибка	95% доверительный интервал	
			Нижняя граница	Верхняя граница
A1	3,802	,564	2,421	5,183
A2	4,743	,564	3,362	6,124
A3	5,200	,564	3,819	6,581

Далее выведены результаты расчета апостериорных критериев, в частности теста Шеффе, по сравнению отдельных типов удобрений и сортов картофеля. По результатам расчета можно видеть, что удобрения В4 и В3 наиболее значимы по своему воздействию на урожайность картофеля и отличаются от других типов удобрений, и что сорт картофеля А3 является наиболее урожайным.

Апостериорные критерии Удобрения

		Множественные сравнения					Однородные подмножества	
Урожайность Шеффе		(I-J)-я разность средних	Стд. Ошибка	Знач.	95% доверительный интервал		Урожайность Шеффе ^{a,b}	
(I) Удобрения	(J) Удобрения				Нижняя граница	Верхняя граница	Удобрения	М
B1	B2	-,02867	,921562	1,000	-3,51007	3,45274	B1	3
	B3	-,49800	,921562	,959	-3,97941	2,98341		3
	B4	-,64367	,921562	,918	-4,12507	2,83774		3
B2	B1	,02867	,921562	1,000	-3,45274	3,51007	B2	3
	B3	-,46933	,921562	,965	-3,95074	3,01207	B3	3
	B4	-,61500	,921562	,927	-4,09641	2,86641	B4	3
B3	B1	,49800	,921562	,959	-2,98341	3,97941	Знач.	
	B2	,46933	,921562	,965	-3,01207	3,95074	,918	
	B4	-,14567	,921562	,999	-3,62707	3,33574		
B4	B1	,64367	,921562	,918	-2,83774	4,12507		
	B2	,61500	,921562	,927	-2,86641	4,09641		
	B3	,14567	,921562	,999	-3,33574	3,62707		

Основанные на наблюдаемых средних.
Член, содержащий ошибку - средний квадрат(Ошибка) = 1,274.

Однородные подмножества

Урожайность Шеффе ^{a,b}		Подмножества	
Удобрения	М	1	
B1	3	4,28900	
B2	3	4,31767	
B3	3	4,78700	
B4	3	4,93267	
Знач.		,918	

Визуализируются средние по группам в однородных наборах. Основанные на наблюдаемых средних.
Член, содержащий ошибку - средний квадрат (Error) = 1,274.
а. Объем выборки, вычисленный как гармоническое среднее, равен 3,000.
b. Альфа = ,05

Сорт

		Множественные сравнения					Однородные подмножества	
Урожайность Шеффе		(I-J)-я разность средних	Стд. Ошибка	Знач.	95% доверительный интервал		Урожайность Шеффе ^{a,b}	
(I) Сорт	(J) Сорт				Нижняя граница	Верхняя граница	Сорт	М
A1	A2	-,94150	,798097	,535	-3,50120	1,61820	A1	4
	A3	-1,39800	,798097	,290	-3,95770	1,16170		4
A2	A1	,94150	,798097	,535	-1,61820	3,50120	A2	4
	A3	-,45650	,798097	,853	-3,01620	2,10320		4
A3	A1	1,39800	,798097	,290	-1,16170	3,95770	A3	4
	A2	,45650	,798097	,853	-2,10320	3,01620		4

Основанные на наблюдаемых средних.
Член, содержащий ошибку - средний квадрат(Ошибка) = 1,274.

Урожайность Шеффе ^{a,b}		Подмножества	
Сорт	М	1	
A1	4	3,80175	
A2	4	4,74325	
A3	4	5,19975	
Знач.		,290	

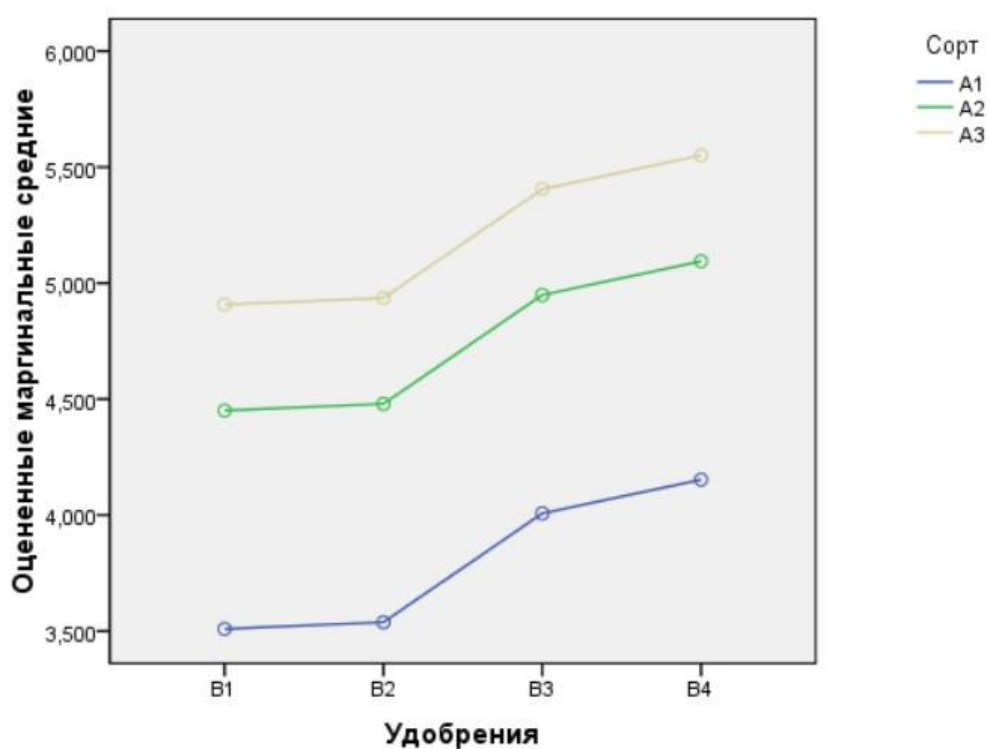
Визуализируются средние по группам в однородных наборах. Основанные на наблюдаемых средних.
Член, содержащий ошибку - средний квадрат (Error) = 1,274.
а. Объем выборки, вычисленный как гармоническое среднее, равен 4,000.
b. Альфа = ,05

Вывод результатов заканчивается графиками профилей, на которых представлены зависимости средней урожайности картофеля от типа удобрений и от сорта.

Из данного графика видно, что наиболее эффективно на среднюю урожайность оказывает влияние удобрения В4 и наименее эффективно – удобрение В1. Если рассматривать влияние сорта, можно заметить, что наиболее урожайным является сорт А3 и наименее урожайным – сорт А1.

Графики профилей

Оцененные маргинальные средние Урожайность



Таким образом, в результате решения данной задачи методом дисперсионного анализа, реализованного в алгоритме одномерной линейной модели, нулевая гипотеза $H_{0,A}$ и нулевая гипотеза $H_{0,B}$ отвергнуты и приняты альтернативные гипотезы о существенном влиянии названного фактора на урожайность.

Вывод: приобрёл практические навыки применения дисперсионного анализа для решения конкретных задач с использованием статистического пакета SPSS.