**Министерство образования и науки Украины**

**Севастопольский национальный технический**

**университет**

## МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к выполнению лабораторной работы № 3 и 4

**” Дисперсионный анализ при помощи системы**

**MINITAB для WINDOWS “**

# **по учебной дисциплине “Прикладная статистика”**

для студентов экономических специальностей

всех форм обучения

**Севастополь**

**2008**

Методические указания рассмотрены и утверждены на заседании кафедры менеджмента и экономико -математических методов протокол № “\_\_\_\_\_” от “\_\_\_\_\_\_\_\_\_\_\_\_\_\_” 2008г.

Рецензент: доцент департамента учета и аудита Т.А.Мараховская

1. **Цель работы**

Изучение возможностей дисперсионного анализа, для выявления зависимостей между экономическими показателями и получение практических навыков работы в системе MINITAB.

**Теоретические сведения**

* 1. **Дисперсионный анализ**
     1. **Однофакторный дисперсионный анализ**

При проведении экономического анализа часто необходимо оценить влияние на целевую функцию y качественного фактора x . Таким фактором могут быть, например, партии сырья, отрасли промышленности, регионы и т.д.

Пусть данные о влиянии некоторого качественного фактора на количественный в форме таблицы.

Таблица 1.1. – влияние качественного фактора на исследуемый показатель

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | … |  |
|  |  | …. |  |
|  |  | … |  |
| … | … | … | … |
|  |  |  |  |

Модель зависимости значений от фактора столбцов можно представить в следующем виде [1-4]:



где - общее среднее, -отклонение от общего среднего для j-го уровня фактора, - случайная составляющая.

По выборочным данным можно вычислить:

1. среднее  для каждого уровня фактора (среднее по столбцам)xj (j=1,2,...u ), по mj параллельным опытам, где mj – число данных в столбце j:

;

1. общее среднее  по всем N опытам, т.е. по всем mj параллельным опытам на всех уровнях фактора xj ():

;

1. общую сумму квадратов отклонений Q0:



1. сумму квадратов, характеризующую влияние фактора x (отклонения между группами)

;

1. остаточную сумму квадратов, зависящую от ошибки ε (отклонения внутри групп)

.

Тождество дисперсионного анализа имеет вид:



На основании вычисленных сумм квадратов вычисляются:

1. оценка дисперсии относительно общего среднего:

,

где  - число степеней свободы;

2) оценка дисперсии «между группами», определяемыми уровнями xj:



где число степеней свободы .

3) выборочная оценка дисперсии «внутри групп», вычисляемая как средняя оценка по всем u группам:



с числом степеней свободы 

Числа степеней свободы должны удовлетворять соотношению



Для того, чтобы сделать вывод о том, влияет ли на исследуемые показатели качественный фактор, сопоставляют дисперсию между группами с общей дисперсией. При этом выдвигают следующие гипотезы:

H0:  , **т.е средние значения по всем столбцам равны** и равны общему среднему, откуда следует, что среднеквадратическое отклонение по факторам равно среднеквадратическому отклонению по всем данным и равно нулю. Т.е. качественный фактор не оказывает влияния на исследуемый показатель.

H1: , , т.е **средние значения по всем столбцам не равны** между собой и не равны общему среднему, откуда следует, что среднеквадратическое отклонение по факторам не совпадает со среднеквадратическим отклонением по всем данным. Т.е. качественный фактор оказывает существенное влияние на исследуемый показатель.

Оценивание значимости влияния фактора x выполняется по F-критерию Фишера, для чего формируется следующее F-отношение:

.

Фактор x признается незначимым, если соответствующее F-отношение оказывается меньше критического, выбранного из таблиц для принятого уровня значимости  и числа степеней свободы сравниваемых дисперсий и .

Табличное значение критерия Фишера определяется дл числа степеней свободы u-1 и N-1 и вероятности ошибки .

Т.е если , то принимается нулевая гипотеза при соответствующем уровне значимости о том, что исследуемый фактор не оказывает существенного влияния на количественные данные.

Если , то нулевая гипотеза отвергается и принимается альтернативная при соответствующем уровне значимости. Исходя из этого, можно сделать вывод о том, что исследуемый фактор оказывает существенное влияние на количественные данные.

Результаты дисперсионного анализа сводятся в таблицу 2.

#### **Таблица 2 Однофакторный дисперсионный анализ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Источник изменчивости** | **Сумма квадратов отклонений** | **Число степеней свободы** | **Оценка дисперсии** | **F – отношение** |
| Между группами |  |  |  |  |
| Внутри групп  ( ошибка ε) |  |  |  |
| Общая сумма |  |  |  |

 - число данных в столбце, u- число столбцов, m – число строк.

**2.1.2. Двухфакторный дисперсионный анализ при перекрестной**

**классификации факторов**

Часто необходимо качественно оценить значимость или незначимость влияния на целевую функцию υ двух одновременно действующих факторов x1 и x2 . Такими факторами могут быть, например, форма собственности предприятия x1 ивид экономической деятельности x2.

Модель двухфакторного дисперсионного анализа имеет вид [1-4]:



где - общее среднее, -отклонение от общего среднего для фактора x1, - отклонение от общего среднего для фактора x2, - отклонение от общего среднего для взаимодействия двух факторов, - случайная составляющая.

В этом случае общую сумму квадратов отклонений Q0 можно разбить на четыре суммы:

1. Qx1-по фактору x1,
2. Qx2-по фактору x2,
3. Qε-остаточную сумму квадратов, зависящую от ошибки ε,
4. Q x1x2-зависящую от взаимодействия (произведения) x1x2 двух факторов.

В этом случае по выборочным значениям вычисляются:

1. среднее  для каждого уровня фактораx1:

;

1. среднее для каждого уровня фактора x2:

;

1. общее среднее  по всем N опытам, т.е. по всем m параллельным опытам на всех сочетаниях уровней факторов x1 и x2 ():

;

1. среднее  по m параллельным опытам для каждого сочетания уровней факторов x1 и x2:

.

В табл.2 показаны данные полного факторного эксперимента с одинаковым числом наблюдений в ячейках.

**Таблица 3. - Данные эксперимента и расчёты средних при двухфакторном дисперсионном анализе**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| j = | | 1 | 2 | … |  |  |
| i = | k |  |  | … |  |  |
| 1 | 1 |  |  |  |  |
| 2 |  |  |  |  |
| … |  | … |  |  |
| m |  |  |  |  |
| .  .  . | 1 |  |  |  |  |  |
| 2 |  |  |  |  |
| … |  |  |  |  |
| m |  |  |  |  |
|  | 1 |  |  |  |  |  |
| 2 |  |  |  |  |
| … |  |  |  |  |
| m |  |  |  |  |
|  |  |  |  |  |  |  |

В табл.2 вычисляется по выделенной части столбца, содержащей m параллельных опытов.

Общая сумма квадратов отклонений Q0 рассчитывается по формуле:



Эту сумму можно разложить на 4 составляющие:

1. сумму, характеризующую влияние фактора x1:

;

1. сумму, характеризующую влияние фактора x2:

;

1. сумму, характеризующую результат влияния взаимодействия x1x2:



1. сумму, характеризующую влияние ошибки ε:



Указанные пять сумм, поделенные на соответствующее число степеней свободы, дают пять различных оценок дисперсии, если влияние факторов x1 и x2 незначимо. Для проведения дисперсионного анализа вычисляются следующие дисперсии:

1. оценка дисперсии относительно общего среднего:

,

где  -общее число наблюдений, а число степеней свободы

;

1. оценка дисперсии «между строками», определяемыми уровнями x1j:

,

где  - число степеней свободы.

1. оценка дисперсии «между столбцами», соответствующими уровням фактора x2:

,

где  - число степеней свободы;

1. оценка дисперсии «между сериями» по m параллельным опытам каждая



с числом степеней свободы ;

1. оценка дисперсии «внутри серий» по m параллельным опытам, вычисляемая как средняя оценка по всем u1u2 сериям:



с числом степеней свободы .

Числа степеней свободы должны удовлетворять соотношению



Статистическое оценивание значимости влияния факторов x1 , x2 и взаимодействия x1x2 выполняются по F-критерию Фишера, для чего формируются следующие F-отношения:

, , .

Фактор x1 или x2 , или взаимодействие x1x2 признаются незначимым, если соответствующее F-отношение оказывается меньше критического, выбранного из таблиц для принятого уровня значимости  и числа степеней свободы сравниваемых дисперсий.

Для того, чтобы сделать вывод о том, влияют ли на исследуемые показатели качественные факторы, выдвигают следующие гипотезы:

H0:  , **т.е средние значения по всем столбцам равны** фактор столбца не оказывает влияния на исследуемый показатель.

H1: , , т.е **средние значения по всем столбцам не равны** фактор столбца оказывает существенное влияние на исследуемый показатель.

H0:  , **т.е средние значения по всем строкам равны** фактор строки не оказывает влияния на исследуемый показатель.

H1: , , т.е **средние значения по всем строкам не равны** фактор строки оказывает существенное влияние на исследуемый показатель.

H0:  , **т.е отклонение взаимодействия факторов равно нулю и взаимодействие не значимо.**.

H1: , фактор взаимодействия значим..

Если , то принимается нулевая гипотеза при соответствующем уровне значимости о том, что исследуемый фактор не оказывает существенного влияния на количественные данные.

Если , то нулевая гипотеза отвергается и принимается альтернативная при соответствующем уровне значимости. Исходя из этого, можно сделать вывод о том, что исследуемый фактор оказывает существенное влияние на количественные данные.

Результаты двухфакторного дисперсионного анализа представляются в виде табл.3.

**Таблица 3. - Двухфакторный дисперсионный анализ при равном числе наблюдений в ячейках**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Вид изменчивости | Сумма квадратов отклонений | Число степеней свободы | Оценка дисперсии | F – отношение |
| От фактора  x1 |  |  |  |  |
| От фактора  x2 |  |  |  |  |
| От взаимо-действия  x1x2 |  |  |  |  |
| Остаточная  (от ε) |  |  |  |  |
| Общая |  |  |  |  |

m – число данных в строке (число повторов в ячейке), - число столбцов, - число строк.

**3. Дисперсионный анализ в системе MINITAB**

Для проведения дисперсионного анализа в системе MINITAB необходимо выбрать из меню  **Stat > ANOVA**.

Различные возможности проведения дисперсионного анализа представлены следующими командами.

Команда **Oneway** позволяет провести однофакторный дисперсионный анализ, если значения выходного и влияющего параметра записаны в двух столбцах.

Команда **Oneway(Unstacked)** позволяет провести однофакторный дисперсионный анализ, если значения выходного параметра разбито на группы и значения для каждой группы записаны в разных столбцах.

Команда **Twoway** позволяет провести двухфакторный анализ для сбалансированных данных (с одинаковым количеством значений в каждой ячейке).

Команда **Balanced ANOVA** позволяет провести многофакторный дисперсионный анализ для сбалансированных моделей с перекрестной и иерархической классификацией.

Команда **General Linear Model** позволяет провести многофакторный несбалансированный дисперсионный анализ для моделей с перекрестной и иерархической классификацией.

# **3.2.1. Однофакторный дисперсионный анализ**

Для проведения однофакторного дисперсионного анализа необходимо подготовить данные в двух столбцах (в первом – входная переменная, качественная, во втором – выходная переменная), выбрать из меню **Stat > ANOVA > Oneway** и заполнить открывшееся диалоговое окно.

Диалоговое окно.

1. Отклик **(Response)** – выберите столбец, содержащий выходную (зависимую) переменную. Столбец должен содержать только числовые значения.
2. Фактор **(Factor)** – выберите столбец, содержащий качественную переменную, влияние которой исследуется. Фактор может иметь как числовые, так и символьные значения.
3. Сохранить остатки **(Store Residuals)**, выбирается, если необходимо сохранить остатки для последующего анализа. Остатки сохраняются в свободном столбце.
4. Сохранить оценки **(Store fits)** Для однофакторного анализа оценки это средние значения для каждого уровня фактора.
5. Графики **<Graphs>** представляют данные в виде точечных и блочных диаграмм для каждой группы с отмеченным средним значением.

### Пример 1

Пусть данные о проценте износа оборудования для 12 предприятий разных отраслей промышленности и форм собственности представлены следующей таблицей.

Таблица 4.

## Исходные данные

|  |  |  |
| --- | --- | --- |
| Field | Owner | d |
| Пищевая | Частн | 31 |
| Пищевая | Частн | 49 |
| Пищевая | Частн | 37 |
| Пищевая | Госуд | 47 |
| Пищевая | Госуд | 57 |
| Пищевая | Госуд | 53 |
| Машиностр | Госуд | 43 |
| Машиностр | Госуд | 59 |
| Машиностр | Госуд | 56 |
| Машиностр | Частн | 47 |
| Машиностр | Частн | 51 |
| Машиностр | Частн | 53 |

Определим зависимость износа оборудования от отрасли промышленности.

В этом случае в диалоговом окне указываются следующие значения

**Response: d**

**Factor: field**

Результаты дисперсионного анализа включают таблицу анализа дисперсии, таблицу средних значений уровней факторов, индивидуальные доверительные интервалы для каждого уровня и общее стандартное отклонение. На рис.1 представлен листинг результатов вычислений. На рисунке используются следующие обозначения:

DF – число степеней свободы,

SS - сумма квадратов,

MS – средний квадрат,

F - отношение Фишера,

P - уровень значимости для вычисленного F,

Level – уровень фактора,

Mean – среднее значение,

StDev – стандартное отклонение.

**One-Way Analysis of Variance**

Analysis of Variance for d

Source DF SS MS F P

field 1 102.1 102.1 1.55 0.241

Error 10 656.8 65.7

Total 11 758.9

Individual 95% CIs For Mean

Based on Pooled StDev

Level N Mean StDev -------+---------+---------+---------

Пищевая 6 45.667 9.852 (-----------\*-----------)

Машиност 6 51.500 5.857 (-----------\*-----------)

-------+---------+---------+---------

Pooled StDev = 8.105 42.0 48.0 54.0

Рис.1 Листинг результатов вычислений для однофакторной модели

Если значения выходной переменной разбито на группы и каждая группа записана в отдельном столбце, то для проведения однофакторного дисперсионного анализа необходимо выбрать из меню **Stat > ANOVA > Oneway [Unstacked]** и заполнить следующее диалоговое окно.

Диалоговое окно

1. Отклик в нескольких столбцах **Responses [in separate columns]** - выберите столбцы, содержащие выходную (зависимую) переменную. Столбцы должны содержать только числовые значения. Система не требует, чтобы в каждом столбце было одинаковое число наблюдений.
2. Графики **<Graphs>** представляют данные в виде точечных и блочных диаграмм для каждой группы с отмеченным средним значением.

### Пример 2

Пусть данные о проценте износа оборудования для 12 предприятий двух отраслей промышленности (пищевая - field1, машиностроение - field2) представлены в табл.5.

Таблица 5.

## Исходные данные

|  |  |
| --- | --- |
| Field1 | Field2 |
| 31 | 59 |
| 49 | 56 |
| 37 | 47 |
| 47 | 51 |
| 57 | 53 |
| 53 |  |
| 43 |  |

В этом случае в диалоговом окне указываются следующие значения.

**Responses [in separate columns]: field1 field2**

Результатом дисперсионного анализа будет таблица представленная на рис.2.

**One-Way Analysis of Variance**

Analysis of Variance

Source DF SS MS F P

Factor 1 182.7 182.7 3.17 0.105

Error 10 576.2 57.6

Total 11 758.9

Individual 95% CIs For Mean

Based on Pooled StDev

Level N Mean StDev ------+---------+---------+---------+

field1 7 45.286 9.050 (---------\*----------)

field2 5 53.200 4.604 (------------\*-----------)

------+---------+---------+---------+

Pooled StDev = 7.591 42.0 48.0 54.0 60.0

Рис.2 Листинг результатов вычислений

Из полученных результатов видно, что P> (=0.05), значит принимается нулевая гипотеза и мы можем сделать вывод о том, что влияние фактора отрасли на уровень износа оборудования незначимо.

Если в опции **<Graphs>** указать **Dotplots of data:√**, то будет построен следующий график (чертой отмечено среднее значение для группы).



Рис.3 Представление экспериментальных данных

# **3.2.2. Двухфакторный дисперсионный анализ**

Для проведения двухфакторного дисперсионного анализа необходимо подготовить данные, выбрать из меню **Stat > ANOVA > Balanced ANOVA** и заполнить открывшееся диалоговое окно.

Эта функция позволяет проводить, как одномерный, так и многомерный анализ дисперсии. Факторы могут быть связаны как перекрестно, так и иерархически, они могут быть детерминированными и случайными, однако данные должны быть сбалансированы. Это значит, что для каждого уровня A должны быть одинаковые уровни фактора B, и в том же количестве.

Диалоговое окно.

1. Отклики **(Responses)** – выберите столбцы, содержащие выходные (зависимые) переменные. Система позволяет анализировать до 50 выходных переменных.
2. Модель **(Model) –** укажите переменные или их комбинацию, которые включаются в модель.
3. Случайные факторы **(Random Factors)** – укажите столбец, содержащий случайную переменную.

### Пример 3

Пусть данные о проценте износа оборудования для 12 предприятий разных отраслей промышленности и форм собственности представлены в табл.1. Определим, как влияют отрасль промышленности, форма собственности и их взаимодействие на процент износа оборудование. Для этого выберем из меню **Stat > ANOVA > Balanced ANOVA**  и заполним диалоговое окно следующим образом

**Responses: d**

**Model: field owner field\*owner**

Результаты дисперсионного анализа представлены на рис.4.

**Analysis of Variance (Balanced Designs)**

Factor Type Levels Values

field fixed 2 Пищевая Машиностр

owner fixed 2 частн госуд

Analysis of Variance for d

Source DF SS MS F P

field 1 102.08 102.08 2.14 0.182

owner 1 184.08 184.08 3.86 0.085

field\*owner 1 90.75 90.75 1.90 0.205

Error 8 382.00 47.75

Total 11 758.92

#### Рис.4 Листинг результатов вычислений для двухфакторной модели

Проанализируем полученные результатs/

Для фактора отрасли P> (=0.05), значит принимается нулевая гипотеза о том, что фактор отрасли не влияет на уровень износа оборудования.

Для фактора формы собственности P> (=0.05), значит принимается нулевая гипотеза о том, что фактор формы собственности не влияет на уровень износа оборудования. Аналогичным образом делаем вывод о том, что на уровень износа оборудование не влияет взаимодействие факторов.

Для анализа многофакторных моделей по несбалансированным данным необходимо выбрать из меню **Stat > ANOVA > General Linear Model.**

4 Выполнение дисперсионного анализа в Excel

Рассмотрим дисперсионный анализ на следующем примере: за месяц известны данные о выработке рабочего за время работы в первую и во вторую смены.

Таблица 2 - Исходные данные

|  |  |
| --- | --- |
| Смена | Выработка рабочего, нормо-час |
| 1 | 12,1; 11,1; 12,6; 12,9; 11,6; 13,1; 12,6; 12,4; 11,6; 17,3; 12,9; 11,6; 12,4 |
| 2 | 9,9; 11,4; 13,4; 10,4; 12,9; 12,6; 13,9; 13,4; 12,4; 9,9; 10,2; 11,2; 9,7 |

Можно ли считать, что расхождение между уровнями выработки рабочего в первую и во вторую смены несущественно, т.е. можно ли считать, что генеральные средние в двух подгруппах одинаковы и, следовательно, выработка рабочего может быть охарактеризована общей средней.

Решение.

Для того чтобы ответить на поставленные вопросы, рассчитаем среднюю выработку рабочих в каждой смене. Величина выработки в первую и вторую смены различна. Теперь возникает вопрос о том, насколько существенны эти расхождения, нужно проверить предположение о возможном влиянии сменности на выработку рабочих. Результаты расчетов сведены в таблицу 3.

Таблица 3 – Промежуточные расчеты для проведения дисперсионного анализа

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Смена | Средняя выработка, нормо-часы | Число смен в месяце | Сумма квадратов отклонений вариантов от групповой средней | Квадраты отклонений групповых средних от общей средней |
| 1 | 12.6308 | 13 | 28.09 | 3,2001 |
| 2 | 11.6385 | 13 | 28.08 | 3,2008 |
| Итого |  | 26 | =56.1585 | =6,4008 |

Используя данные таблицы, рассчитаем  и .

Число степеней свободы для расчета внутригрупповой дисперсии равно () 24 (26-2), а для расчета межгрупповой дисперсии число степеней свободы равно  - 1 (2-1).





Рассчитаем значение критерия Фишера по следующей формуле:

 (4)



В соответствии с числом степеней свободы для расчета внутригрупповой и межгрупповой дисперсий (24 и 1) в таблице F-распределения для α=5% находим Fтабл = 4.26.

При этом выдвигается две гипотезы. Нулевая гипотеза гласит о том, что различия выработки рабочего в первую и вторую смены несущественны. Альтернативная гипотеза: существуют существенные различия в значении выработки рабочего в первую и во вторую смены.

Так как расчетное значение критерия Фишера значительно меньше табличного значения критерия Фишера, то гипотеза о несущественности различия выработки рабочего в первую и вторую смены не опровергается, т.е. сменность не оказывает влияния на уровень выработки рабочего.

Для того, чтобы провести дисперсионный анализ в Excel, необходимо активировать команду «Анализ данных». Для этого проходится следующий путь: Сервис -> Надстройки -> Пакет анализа. После этого в меню «Сервис» появляется команда «Анализ данных» и выбирается команда «Однофакторный дисперсионный анализ».

Далее необходимо заполнить окно «Однофакторный дисперсионный анализ»:

«Входной интервал» - вводится ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.

«Группирование» - установите переключатель в положение. По столбцам или По строкам в зависимости от расположения данных во входном диапазоне.

«Метки в первой строке/Метки в первом столбце» - если первая строка исходного диапазона содержит названия столбцов, установите переключатель в положение Метки в первой строке. Если названия строк находятся в первом столбце входного диапазона, установите переключатель в положение Метки в первом столбце. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

«Альфа» - введите уровень значимости, необходимый для оценки критических параметров F-статистики. Уровень альфа связан с вероятностью возникновения ошибки типа I (опровержение верной гипотезы).

«Выходной диапазон» - введите ссылку на левую верхнюю ячейку выходного диапазона. Размеры выходной области будут рассчитаны автоматически, и соответствующее сообщение появится на экране в том случае, если выходной диапазон занимает место существующих данных или его размеры превышают размеры листа.

«Новый лист» - установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

«Новая книга» - установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку A1 на первом листе в этой книге.

### Пример заполнения окна «Однофакторный дисперсионный анализ» представлен на рисунке 2.

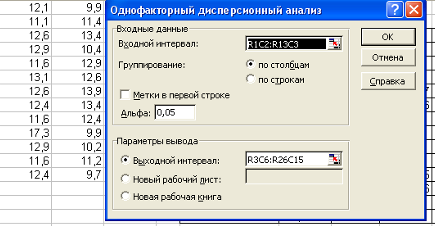


Рисунок 2 – Пример заполнения окна «Однофакторный дисперсионный анализ»

Результаты расчетов однофакторного дисперсионного анализа представлены на рисунке 3.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Однофакторный дисперсионный анализ |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ИТОГИ |  |  |  |  |  |  |
| Группы | Счет | Сумма | Среднее | Дисперсия |  |  |
| Столбец 1 | 13 | 164,2 | 12,63077 | 2,34064103 |  |  |
| Столбец 2 | 13 | 151,3 | 11,63846 | 2,33923077 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Дисперсионный анализ |  |  |  |  |  |  |
| Источник вариации | SS | df | MS | F | P-Значение | F критическое |
| Между группами | 6,400385 | 1 | 6,400385 | 2,73528203 | 0,111176312 | 4,259675279 |
| Внутри групп | 56,15846 | 24 | 2,339936 |  |  |  |
|  |  |  |  |  |  |  |
| Итого | 62,55885 | 25 |  |  |  |  |

Рисунок 3 – Результаты расчетов по однофакторному дисперсионному анализу

Интерпретация результатов:

«Группы» - данные по выработке в первую и вторую смены.

«Счет» - количество наблюдений в каждой из групп.

«Сумма» - сумма элементов каждой из групп.

«Среднее» - средняя выработка в каждой из групп.

«Дисперсия» - рассчитывается дисперсия по каждой из групп;

SS - сумма квадратов;

df - число степеней свободы;

MS – средний квадрат;

F – расчетное значение отношения Фишера;

P - уровень значимости для вычисленного F;

F критическое – табличное значение отношения Фишера.

Результаты расчетов аналогичны результатам, полученным при расчетах вручную.

**Двухфакторный дисперсионный анализ в MS Exel**

Используя данный предыдущего примера, предположим, что у нас есть данные о поле работников. Для проведения двухфакторного дисперсионного анализа в MS Exel необходимо представить данные в виде перекрестной классификации:

|  |  |  |
| --- | --- | --- |
|  | 1 | 2 |
| муж | 12,1 | 9,9 |
| 11,1 | 11,4 |
| 12,6 | 13,4 |
| 12,9 | 10,4 |
| 11,6 | 12,9 |
| 13,1 | 12,6 |
| 12,6 | 13,9 |
| жен | 12,4 | 13,4 |
| 11,6 | 12,4 |
| 17,3 | 9,9 |
| 12,9 | 10,2 |
| 11,6 | 11,2 |
| 12,4 | 9,7 |
| 13,1 | 12,6 |

В меню «Сервис» выбрать команду «Анализ данных» и команду «Двухфакторный дисперсионный анализ с повторениями».

Далее необходимо заполнить окно «Двухфакторный дисперсионный анализ с повторениями»:

«Входной интервал» - вводится ссылка на диапазон, содержащий анализируемые данные.Необходимо отметить не только сами числа, но и заголовок таблицы.

«Число строк для выборки» - необходимо ввести количество повторений в одной ячейке. (Для нашего примера - 7)

«Альфа» - введите уровень значимости, необходимый для оценки критических параметров F-статистики. Уровень альфа связан с вероятностью возникновения ошибки типа I (опровержение верной гипотезы).

«Выходной диапазон» - введите ссылку на левую верхнюю ячейку выходного диапазона. Размеры выходной области будут рассчитаны автоматически, и соответствующее сообщение появится на экране в том случае, если выходной диапазон занимает место существующих данных или его размеры превышают размеры листа.

«Новый лист» - установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

«Новая книга» - установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку A1 на первом листе в этой книге.

Пример заполнения окна «Однофакторный дисперсионный анализ» представлен на рисунке 2.

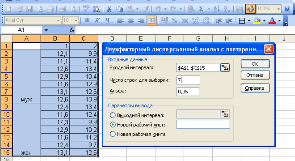


Рисунок 2 – Пример заполнения окна «Двухфакторный дисперсионный анализ»

Результаты расчетов двухфакторного дисперсионного анализа представлены на рисунке 3.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Дисперсионный анализ |  |  |  |  |  |  |
| Источник вариации | SS | df | MS | F | P-Значение | F критическое |
| Выборка | 0,001429 | 1 | 0,001429 | 0,000643 | 0,979986 | 4,259677 |
| Столбцы | 6,412857 | 1 | 6,412857 | 2,884498 | 0,102366 | 4,259677 |
| Взаимодействие | 3,862857 | 1 | 3,862857 | 1,73751 | 0,199898 | 4,259677 |
| Внутри | 53,35714 | 24 | 2,223214 |  |  |  |
|  |  |  |  |  |  |  |
| Итого | 63,63429 | 27 |  |  |  |  |

Рисунок 3 – Результаты расчетов по однофакторному дисперсионному анализу

Интерпретация результатов:

SS - сумма квадратов;

df - число степеней свободы;

MS – средний квадрат;

F – расчетное значение отношения Фишера;

P - уровень значимости для вычисленного F;

F критическое – табличное значение отношения Фишера.

1. **Задание по выполнению лабораторной работы**
   1. **Однофакторный дисперсионный анализ**

Вы собираетесь открывать магазин одежды. Произведенный опрос среди предполагаемых покупателей позволил получить вам примерный уровень доходов респондентов в месяц, которые предпочитают одежду тех или иных торговых марок. Необходимо проверить, есть ли существенное различие в уровне доходов и маркой одежды, которую предпочитают покупатели. Выясните, какие торговые марки можно отнести к одной группе (по величине объема продаж) и предположите, как их можно сегментировать.

В табл.6 приведены варианты заданий.

Таблица 6.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Торговые марки | | | | | | | | | | | | | | | | | | | |
| M1 | M2 | | M3 | | M4 | | M5 | | M6 | | M7 | | M8 | | M9 | | M10 | M11 | M12 |
| 555 | 1810 | | 1749 | | 2711 | | 994 | | 3687 | | 566 | | 4691 | | 1679 | | 861 | 1446 | 3543 |
| 426 | 1122 | | 1746 | | 2514 | | 1085 | | 2489 | | 883 | | 4130 | | 2838 | | 1074 | 1010 | 4828 |
| 349 | 2220 | | 1509 | | 2177 | | 1215 | | 2717 | | 844 | | 5328 | | 3615 | | 920 | 1414 | 5027 |
| 506 | 720 | | 1949 | | 2754 | | 1024 | | 4055 | | 917 | | 3268 | | 2098 | | 1192 | 1528 | 2937 |
| 550 | 2347 | | 1673 | | 2482 | | 931 | | 2485 | | 850 | | 3821 | | 2602 | | 970 | 1572 | 3067 |
| 443 | 1841 | | 1275 | | 2219 | | 1242 | | 2322 | | 768 | | 4132 | | 2304 | | 963 | 1538 | 4301 |
| 626 | 2250 | | 1651 | | 3065 | | 948 | | 3548 | | 907 | | 6429 | | 2529 | | 1417 | 1697 | -393 |
| 582 | 2293 | | 1745 | | 2411 | | 1041 | | 3139 | | 983 | | 5833 | | 2531 | | 535 | 1223 | 1687 |
| 463 | 2550 | | 862 | | 2169 | | 948 | | 2258 | | 855 | | 3356 | | 2784 | | 1101 | 1072 | 3623 |
| 306 | 2977 | | 831 | | 2338 | | 976 | | 3327 | | 794 | | 2694 | | 3646 | | 1031 | 1725 | 3187 |
| 566 | 1542 | | 1533 | | 2415 | | 998 | | 2994 | | 815 | | 5074 | | 4089 | | 1011 | 1807 | 3353 |
| 569 | 3322 | | 1432 | | 2255 | | 724 | | 3783 | | 760 | | 3363 | | 2603 | | 1044 | 1512 | 4048 |
| 463 | 1441 | | 1465 | | 2527 | | 952 | | 3996 | | 830 | | 4852 | | 2861 | | 724 | 1623 | 3776 |
| 304 | 1952 | | 1934 | | 2446 | | 998 | | 3199 | | 900 | | 3316 | | 2784 | | 1327 | 1155 | 5251 |
| 528 | 1813 | | 1813 | | 2806 | | 1115 | | 4875 | | 832 | | 1985 | | 2569 | | 1199 | 1200 | 2009 |
| 496 | 617 | | 1744 | | 2618 | | 834 | | 2230 | | 711 | | 4547 | | 3584 | | 1206 | 1302 | 3480 |
| 648 | 2615 | | 1151 | | 2430 | | 1034 | | 3101 | | 797 | | 3293 | | 2153 | | 601 | 1304 | 4627 |
| 457 | 1777 | | 876 | | 2748 | | 1018 | | 4146 | | 936 | | 3922 | | 3421 | | 871 | 1687 | 2355 |
| 690 | 1420 | | 1382 | | 3110 | | 1000 | | 733 | | 809 | | 3086 | | 4068 | | 901 | 1428 | 2329 |
| 548 | 1843 | | 1555 | | 2996 | | 834 | | 3227 | | 729 | | 2447 | | 3080 | | 898 | 1433 | 3920 |
| 491 | 2574 | | 940 | | 2707 | | 1165 | | 2734 | | 926 | | 3524 | | 2831 | | 789 | 1440 | 1922 |
| Вариант | | Торговые марки | | | | | | | | | | | |
| 1 | | M1 | | M2 | | M3 | | M4 | | M5 | | M6 | |
| 2 | | M2 | | M3 | | M4 | | M5 | | M6 | | M7 | |
| 3 | | M3 | | M4 | | M5 | | M6 | | M7 | | M8 | |
| 4 | | M4 | | M5 | | M6 | | M7 | | M8 | | M9 | |
| 5 | | M5 | | M6 | | M7 | | M8 | | M9 | | M10 | |
| 6 | | M1 | | M3 | | M4 | | M5 | | M9 | | M10 | |
| 7 | | M1 | | M4 | | M5 | | M6 | | M9 | | M10 | |
| 8 | | M1 | | M5 | | M6 | | M7 | | M9 | | M10 | |
| 9 | | M1 | | M6 | | M7 | | M8 | | M9 | | M10 | |
| 10 | | M1 | | M3 | | M5 | | M7 | | M9 | | M11 | |
| 11 | | M2 | | M4 | | M5 | | M6 | | М11 | | М12 | |
| 12 | | M2 | | M5 | | M6 | | M7 | | М11 | | М12 | |
| 13 | | M2 | | M6 | | M7 | | M8 | | M10 | | M12 | |
| 14 | | M2 | | M4 | | M6 | | M8 | | M10 | | M12 | |
| 15 | | M2 | | M5 | | M7 | | M8 | | М11 | | М12 | |

**4.2 Двухфакторный дисперсионный анализ**

В таблице приведены данные опроса 32 человек. Опрашиваемые были выбраны случайным образом из групп людей, которые формировались так, чтобы результаты опроса были сбалансированы по всем уровням факторов.

Таблица 7

**Результаты опроса**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Образование | Сфера деятельн. | Пол | Положение | Доход | Расход |
| X1 | X2 | X3 | X4 | Y1 | Y2 |
| Экономич. | Финансы | Муж. | Руковод. | 852 | 650 |
| Экономич. | Финансы | Жен. | Руковод. | 750 | 700 |
| Экономич. | Производ. | Муж. | Руковод. | 210 | 140 |
| Экономич. | Производ. | Жен. | Руковод. | 180 | 160 |
| Экономич. | Сельск,х. | Муж. | Работник | 120 | 80 |
| Экономич. | Сельск,х. | Жен. | Работник | 130 | 120 |
| Экономич. | Образов. | Муж. | Работник | 210 | 180 |
| Экономич. | Образов. | Жен. | Работник | 190 | 170 |
| Технич. | Финансы | Муж. | Работник | 320 | 240 |
| Технич. | Финансы | Жен. | Работник | 240 | 220 |
| Технич. | Производ. | Муж. | Работник | 230 | 180 |
| Технич. | Производ. | Жен. | Работник | 140 | 130 |
| Технич. | Сельск,х. | Муж. | Руковод. | 350 | 300 |
| Технич. | Сельск,х. | Жен. | Руковод. | 360 | 320 |
| Технич. | Образов. | Муж. | Руковод. | 310 | 250 |
| Технич. | Образов. | Жен. | Руковод. | 310 | 300 |
| Медицин, | Финансы | Муж. | Руковод. | 540 | 450 |
| Медицин, | Финансы | Жен. | Руковод. | 450 | 420 |
| Медицин, | Производ. | Муж. | Руковод. | 310 | 210 |
| Медицин, | Производ. | Жен. | Руковод. | 405 | 380 |
| Медицин, | Сельск,х. | Муж. | Работник | 110 | 100 |
| Медицин, | Сельск,х. | Жен. | Работник | 120 | 110 |
| Медицин, | Образов. | Муж. | Работник | 210 | 180 |
| Медицин, | Образов. | Жен. | Работник | 180 | 170 |
| Гуманит. | Финансы | Муж. | Работник | 230 | 160 |
| Гуманит. | Финансы | Жен. | Работник | 240 | 220 |
| Гуманит. | Производ. | Муж. | Работник | 120 | 110 |
| Гуманит. | Производ. | Жен. | Работник | 125 | 120 |
| Гуманит. | Сельск,х. | Муж. | Руковод. | 280 | 180 |
| Гуманит. | Сельск,х. | Жен. | Руковод. | 300 | 280 |
| Гуманит. | Образов. | Муж. | Руковод. | 240 | 230 |
| Гуманит. | Образов. | Жен. | Руковод. | 230 | 200 |

Требуется методом двухфакторного дисперсионного анализа оценить степень влияния изучаемых факторов на результирующий экономический показатель. Первоначально оценить модель без взаимодействия факторов, затем с взаимодействием. Сравнить результаты. Сделать выводы. Варианты заданий приведены в табл.8.

Таблица 8

Варианты заданий

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Вариант | Первый фактор | Второй фактор | Отклик | Вариант | Первый фактор | Второй фактор | Отклик |
| 1 | X1 | X2 | Y1 | 7 | X1 | X2 | Y2 |
| 2 | X1 | X3 | Y1 | 8 | X1 | X3 | Y2 |
| 3 | X1 | X4 | Y1 | 9 | X1 | X4 | Y2 |
| 4 | X2 | X3 | Y1 | 10 | X2 | X3 | Y2 |
| 5 | X2 | X4 | Y1 | 11 | X2 | X4 | Y2 |
| 6 | X3 | X4 | Y1 | 12 | X3 | X4 | Y2 |

1. **Порядок выполнения работы**
2. В соответствии с вариантом задания выполнить однофакторный дисперсионный анализ, сделать выводы, написать отчет.
3. В соответствии с вариантом задания выполнить двухфакторный дисперсионный анализ, сделать выводы, написать отчет.

**Контрольные вопросы**

1. Сформулируйте основную идею дисперсионного анализа, для решения каких задач он наиболее эффективен ?
2. Что показывает F отношение Фишера?
3. Каковы основные теоретические предпосылки дисперсионный анализ?
4. Произведите разложение общей суммы квадратов отклонений на составляющие в однофакторном дисперсионном анализе.
5. Как получить оценки дисперсий из сумм квадратов отклонений? Как получаются необходимые числа степеней свободы?
6. Приведите свой пример двухфакторного дисперсионного анализа.
7. На какие суммы разлагается общая сумма квадратов отклонений в двухфакторном дисперсионном анализе?
8. Поясните схему двухфакторного дисперсионного анализа.
9. Чем отличается перекрестная классификация от иерархической классификации?
10. Чем отличаются сбалансированные данные?

**Литература**

1. Шеффе Г. Дисперсионный анализ. – М.: Наука. 1980.- 512с.
2. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента. Пер. с англ. – М.: Мир, 1981.-520с.
3. Дэниел К. Применение статистики в промышленном эксперименте.-М.:Мир, 1979.-300с.
4. Хикс Ч. Основные принципы планирования эксперимента.- М.:Мир, 1967.

Методические указания разработали: профессор, д.т.н. Цуканов А.В. и к.т.н., доцент, Русина Н.А.