Проверка однородности двух выборок производится с помощью критерия Стьюдента (или *t* – критерия). Рассмотрим постановку задачи проверки однородности двух выборок. Пусть произведено две выборки объёмом http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image564.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image566.gif. Необходимо проверить нулевую гипотезу о том, что генеральные средние двух выборок равны. То есть, http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image568.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image570.gif. *n1*

Прежде чем рассматривать методику решения задачи рассмотрим некоторые теоретические положения, используемые для решения задачи. Известный математик У.С. Госсет (ряд своих работ публиковал под псевдонимом Стьюдент) доказал, что статистика *t* (6.4) подчиняется определенному закону распределения, который в последствии был назван законом распределения Стьюдента (второе название закона – ”*t* – распределение”).

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image572.gif, (6.4)

где

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image574.gif- среднее значение случайной величины *X*;

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image576.gif- математическое ожидание случайной величины *X*;

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image578.gif- среднеквадратичного отклонения среднего http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image574.gifвыборки объёма *n*.

Оценка среднеквадратичного отклонения среднего http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image574.gifрассчитывается по формуле (6.5):

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image580.gif, (6.5)

где

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image582.gif- среднеквадратичного отклонения случайной величины *X*.

Распределение Стьюдента имеет один параметр – количество степеней свободы http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image584.gif.

Теперь вернемся к исходной постановке задачи с двумя выборками и рассмотрим случайную величину равную разности средних двух выборок (6.6):

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image586.gif (6.6)

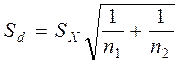
При условии выполнения гипотезы о равенстве генеральных средних справедливо (6.7):

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image588.gif (6.7)

Перепишем соотношение (6.4) применительно нашему случаю:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image590.gif (6.8)

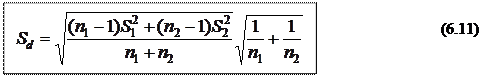
Оценка среднеквадратичного отклонения http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image592.gifдолжна быть выражена через оценку среднеквадратичного отклонения объединенной совокупности http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image594.gif(6.9):

 (6.9)

Оценка дисперсии объединенной совокупности http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image598.gifдолжна быть выражена через оценки дисперсии, рассчитанные по двум выборкам http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image600.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image602.gif:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image604.gif (6.10)

С учетом формулы (6.10) соотношение (6.9) можно переписать в виде (6.11). Соотношение (6.9) является основной расчетной формулой задачи сравнения средних:



При подстановке значения http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image592.gifв формулу (6.8) будем иметь выборочное значение *t*-критерия http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image609.gif. По таблицам распределения Стьюдента при количестве степеней свободы http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image611.gifи заданном уровне значимости http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image487.gifможно определить http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image614.gif. Теперь, в случае если http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image616.gif, то гипотеза о равенстве двух средних отвергается.

Рассмотрим пример выполнения расчетов для проверки гипотезы равенства двух средних в EXCEL. Сформируем таблицу данных (рис. 6.22). Данные сгенерируем с помощью программы генерации случайных чисел пакета ”Анализ данных”:

- X1 выборка из нормального распределения с параметрами http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image618.gifhttp://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image620.gifобъемом http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image622.gif;

- X2 выборка из нормального распределения с параметрами http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image624.gifобъемом http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image626.gif;

- X3 выборка из нормального распределения с параметрами http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image628.gifобъемом http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image630.gif;

- X4 выборка из нормального распределения с параметрами http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image632.gifобъемом http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image634.gif.

Проверим гипотезу равенства двух средних (X1-X2), (X1-X3), (X1-X4). В начале рассчитаем параметры выборок признаков X1-X4 (рис. 6.23). Далее рассчитаем значение *t*- критерия. Расчеты выполнит с помощью формул (6.6) – (6.9) в EXCEL. Результаты расчетов сведем в таблицу (рис. 6.24).



Рис. 6.22. Таблица данных



Рис. 6.23. Параметры выборок признаков X1-X4



Рис. 6.24. Сводная таблица расчета значений *t*– критерия для пар признаков (X1-X2), (X1-X3), (X1-X4)

По результатам, приведенным в таблице на рис. 6.24 можно сделать заключение, что для пары признаков (X1-X2) гипотеза равенства средних двух признаков отвергается, а для пар признаков (X1-X3), (X1-X4) гипотезу можно считать справедливой.

Такие же результаты можно получить с помощью программы “Двухвыборочный *t*-тест с одинаковыми дисперсиями” пакета Анализ данных. Интерфейс программы приведен на рис. 6.25.

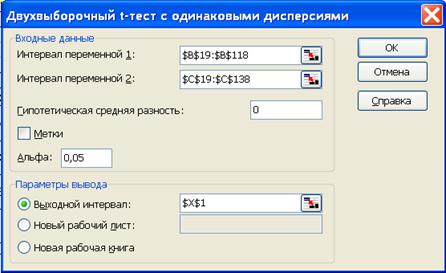


Рис. 6.25. Параметры программы “Двухвыборочный *t*- тест с одинаковыми дисперсиями”

Результаты расчетов проверки гипотез равенства двух средних пар признаков (X1-X2), (X1-X3), (X1-X4), полученные с помощью программы приведены на рис. 6.26-6.28.



Рис. 6.26. Расчет значения *t*– критерия для пары признаков (X1-X2)



Рис. 6.27. Расчет значения *t*– критерия для пары признаков (X1-X3)

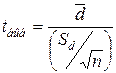


Рис. 6.28. Расчет значения *t*– критерия для пары признаков (X1-X4)

Двухвыборочный *t*-тест с одинаковыми дисперсиями иначе принято называть *t*-тестом с независимыми выборками. Большое распространение так же получил *t*-тестом с зависимыми выборками. Ситуация, когда крайне важно применять данный критерий возникает тогда, когда одна и та же случайная величина подвергается измерению дважды. Количество наблюдений в обоих случаях одинаково. Введем обозначения для двух последовательных измерений некоторого свойства одних и тех же объектовhttp://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image650.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image652.gifhttp://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image654.gif, http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image656.gif, а разность двух последовательных измерений обозначим http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image658.gif:

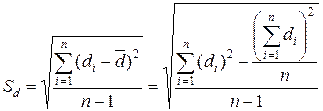
http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image660.gif (6.12)

В этом случае формула для выборочного значения критерия приобретает вид:

, (6.13)

где

 (6.14)

 (6.15)

В этом случае количество степеней свободы http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image668.gif. Проверку гипотезы можно выполнить с помощью программы “Парный двухвыборочный *t*-тест” пакета анализа данных (рис. 6.29).

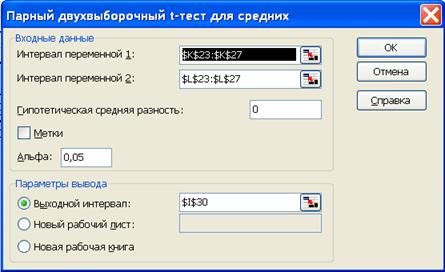


Рис. 6.29. Параметры программы “Парный двухвыборочный *t*-тест”

**6.5. Дисперсионный анализ –классификация по одному признаку (F - критерий)**

В дисперсионном анализе проверяется гипотеза, которая является обобщением гипотезы равенства двух средних на случай, когда проверяется гипотеза равенства одновременно нескольких средних. В дисперсионном анализе исследуется степень влияния одного или нескольких факторных признаков на результативный признак. Идея дисперсионного анализа принадлежит Р. Фишеру. Он использовал его для обработки результатов агрономических опытов. Дисперсионный анализ применяется для установления существенности влияния качественных факторов на исследуемую величину. Английское сокращенное название дисперсионного анализа – ANOVA (analysis variation).

Общая форма представления данных с классификацией по одному признаку представлена в таблице 6.1.

Таблица 6.1. Форма представления данных с классификацией по одному признаку

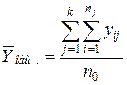
|  |  |
| --- | --- |
|  | Классификация по одному признаку |
| Номер наблюдения | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image672.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image674.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image676.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image680.gif |
|  | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image682.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image684.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image686.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image688.gif |
|  | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image690.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image692.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image694.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image696.gif |
|  | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image698.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image700.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image702.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image704.gif |
| http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif |
| http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif |
| n | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image707.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image709.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image711.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image713.gif |
| Среднее по столбцу | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image715.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image717.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image719.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image721.gif |
| Общее среднее |  |  |  | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image678.gif | http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image723.gif |

Каждый набор значений по столбцам соответствует набору значений одного и того же признака, соответствующего определенному правилу классификации. То есть такая таблица должна быть получена по таблице данных, содержащей два столбца. В первом столбце исходной таблицы располагаются значения исследуемого признака во втором – номера классов заданной классификации. На практике количество наблюдений по столбцам совпадает далеко не всегда. По этой причине при рассмотрении расчетных формул мы будем рассматривать более общий случай, считая, что количество значений в столбцах различно и равно http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image725.gif, где k – количество классов в классификации.

Средние по столбцам рассчитаем по формуле:

 (6.16)

Тогда, учитывая, что общее число наблюдений во всех столбцах равно http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image729.gifhttp://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image731.gif, общее среднее рассчитаем по формуле:

 (6.17)

Для любого наблюдения имеет место следующее равенство:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image735.gif (6.18)

Соотношение (6.18) показывает, что отклонение наблюдения от общего среднего должна быть разделено на две части. Первая часть это отклонение среднего по столбцу от общего среднего (отклонение по столбцу) или факторное отклонение. Вторая часть - ϶ᴛᴏ отклонение наблюдения от среднего по столбцу (остаточное отклонение). Возведем отклонения в формуле (6.18) в квадрат и просуммируем значения по всем наблюдениям. Тогда получим меры вариации:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image737.gif (6.19)

В словесной формулировке эту формулу можно выразить как:

общая вариация = вариация столбцов + остаточная вариация.

Это соотношение принято называть основным тождеством вариации: оно разделяет общую вариацию наблюдений на вариацию, обусловленную классификацией столбца, и вариацию, обусловленную случайной ошибкой.

Проиллюстрируем расчёты вариации на конкретном примере. На рис 6.29 представлена таблица, в которой приведены пятилетние нормы прибыли на акции для фирм в трех отраслях промышленности.

Предварительно расчетным путем убедимся в справедливости соотношения 6.18. Для этого рассчитаем правую и левую часть выражения 6.18 для всех элементов таблицы и проверим тождество. На рис. 30 представлены расчёты для первого столбца таблицы рис. 6.29. Справедливость формулы для остальных столбцов таблицы данных студенты могут проверить самостоятельно, проведя аналогичные расчёты для этих столбцов.

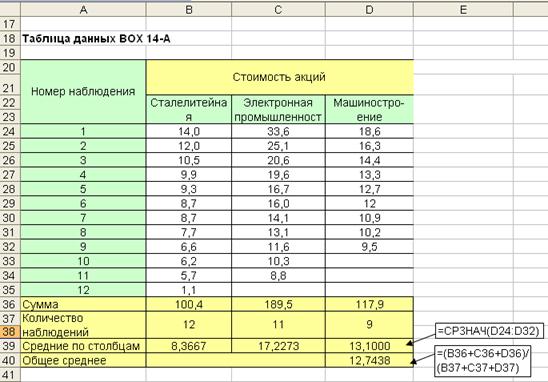


Рис. 6.29. Пятилетние нормы прибыли на акции для фирм в трех отраслях промышленности

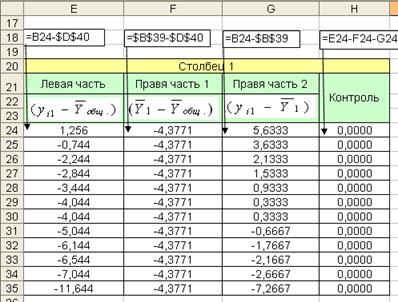


Рис. 6.30. Результаты проверки тождества 6.18

Общая вариация для рассматриваемого примера:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image743.gif (6.20)

Вариацию столбцов рассчитаем по формуле:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image745.gif (6.21)

Остаточная вариация должна быть рассчитана, как по общей формуле, так и как разность общей вариации и вариации столбцов:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image747.gif (6.22)

Контроль вычислений подтверждает справедливость формулы (6.19):

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image749.gif (6.23)

Рассмотрим технику выполнения расчета вариаций в EXCEL. В начале выполним предварительные расчёты, которые приведены на рис. 6.31.

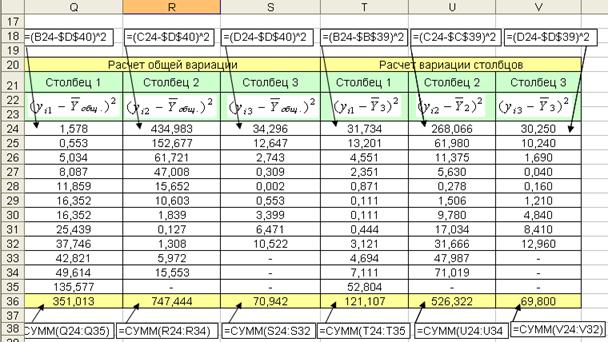


Рис. 6.31. Предварительные расчёты

На базе предварительного расчета произведем окончательный расчет вариаций Рис. 6.32.

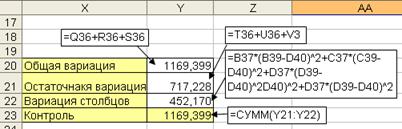


Рис. 6.32. Окончательный расчет вариаций

Статистическая гипотеза состоит в том, что все средние по столбцам равны некоторому общему среднему. В случае если гипотеза справедлива, то вариация столбцов http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image755.gifбудет достаточно мала по сравнению с остаточной вариацией http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image757.gif. Гипотеза проверяется с помощью критерия Фишера:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image759.gif (6.24)

В формуле 6.24 вариация делится на количество степеней свободы (использованных связей). Так как http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image761.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image757.gifимеют соответственно http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image763.gifи http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image765.gifстепеней свободы, то http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image767.gifимеет http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image769.gifстепеней свободы. Стоит сказать, что для нашего примера выборочное значение F - критерия будет равно 9,15:

http://referatwork.ru/image.php?way=oplibru/baza2/2723675952107.files/image771.gif (6.25)

По таблице распределения Фишера при уровне значимости 0,05 и степенях свободы 2 и 29 определяем критическое значения критерия Фишера – 3,32. В EXCEL критическое значение определяем с помощью функции “FРАСПОБР”. Пример обращения к функции приведен на рис. 6.32. Поскольку выборочное значение критерия превосходит критическое, гипотеза отвергается.

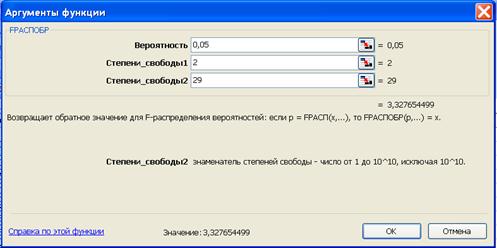


Рис. 6.32. Расчет критического значения для F – распределения.

Расчет параметров необходимых для решения задачи дисперсионного анализа приведен на рис. 6.32.

Для выполнения расчетов, необходимых для проверки гипотезы, выдвигаемой в однофакторном дисперсионном анализе можно воспользоваться специальной программой "Однофакторный дисперсионный анализ" Пакета анализа данных (рис. 6.34). Ввод параметров программы осуществляется в диалоговом окне, представленном на рис. 6.35. Результаты расчетов программы представлены на рис. 6.36. Сравнение результатов по формулам и с помощью программы подтверждают полную идентичность результатов.

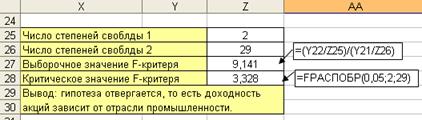


Рис.6.32. Выводы по решению задачи дисперсионного анализа

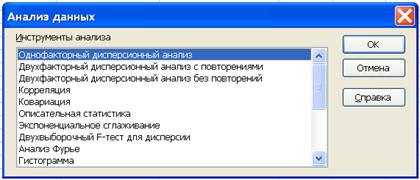


Рис.6.33. Обращение к программе "Однофакторный дисперсионный анализ"

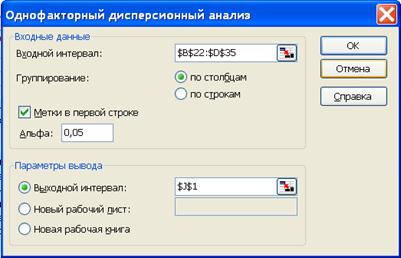


Рис. 6.34. Параметры программы "Однофакторный дисперсионный анализ"

При этом, нельзя не отметить некоторые недостатки работы программы. Проблема состоит в форме представления данных в виде классифицированной таблицы. На практике исходные данные чаще всего представляются в форме таблицы ”объект-свойство”. Такая таблица включает множество признаков, которые используются для решения целого спектра задач анализа данных. Задача дисперсионного анализа должна быть только одной из ряда решаемых задач. Вместе с тем, даже для задачи дисперсионного анализа должна быть определено несколько вариантов классификации. В этом случае для решения частной задачи придется строить вспомогательные таблицы, что крайне неудобно. По этой причине рассмотрим решение задачи дисперсионного анализа в случае представления данных в форме таблицы ”объект-свойство”. Для демонстрации вычислений в EXCEL будем использовать тот же пример с прибыльностью акций по отраслям промышленности. Форма таблицы ”объект-свойство” приведена на рис. 6.36

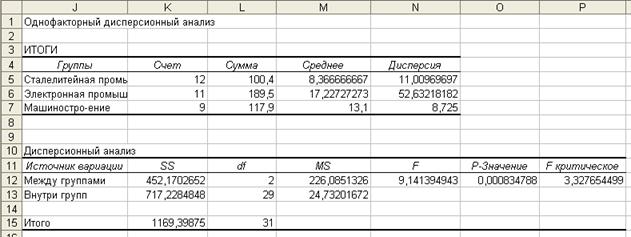


Рис. 6.35. Результаты работы программы "Однофакторный дисперсионный анализ"

Все данные по исследуемому признаку представлены в одном столбце ”Стоимость акций”. Классификация по отраслям задана в отдельном столбце и определена номером класса. Причем, классификация могла быть задана и в текстовом виде. Порядок объектов в данном случае не имеет значения и ни как не отражается на расчетах.

На первом шаге вычислений производится расчет средних значений и определяются входные параметры задачи (рис. 6.36). Для проведения аналогии в таблице на рис. 6.36 используется термин не класс, столбец. Особое внимание при выполнении расчетов крайне важно уделить использованию функции ”СУММЕСЛИ”. Как показал опыт, студенты плохо владеют этой полезной функцией EXCEL. Диалоговое окно функции ”СУММЕСЛИ” приведено на рис. 6.37.



Рис. 6.36. Первый шаг расчета

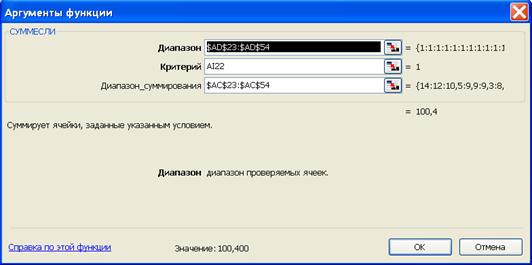


Рис. 6.37. Диалоговое окно функции ”СУММЕСЛИ”

На втором шаге выполняются вспомогательные расчёты в столбцах “Расчет общей вариации” и “Расчет остаточной вариации”. Расчетные формулы приведены на рис. 6.38. Все результаты по расчету параметров необходимых для решения задачи дисперсионного анализа приведены на рис. 6.39.

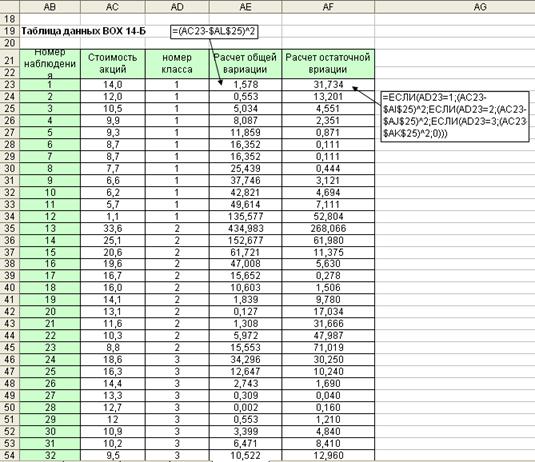


Рис. 6.38. Промежуточные расчёты в таблице ”объект-свойство”



Рис. 6.39. Решение задачи дисперсионного анализа с данными, заданными в таблице ”объект-свойство”