**Лабораторная работа № 12\_1**

**Ковариационный анализ**

Перед выполнением работы необходима консультация с преподавателем.

**I.** В таблице 9 содержатся данные о цене продажи жилых домов.

Табл.9 цена продажи (в дол.), площадь (в квадратных футах)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Категория | | | | | |
| Неплохой (1) | | Хороший (2) | | Превосходный (3) | |
| цена | полезная  площадь | цена | полезная  площадь | цена | полезная  площадь |
| 106000  107500  75500  88000  83000  52980 | 843  907  824  672  698  723 | 146500  128500  179000  219770  139000  131000  197500  116000 | 1317  1080  1688  1738  1040  1404  1558  976 | 166500  275000  164500  333000  330000  302120  229000  201800 | 1318  1507  1292  2350  1732  2010  1749  1762 |
| Среднее  85496,6 | 777,83 | 157158,8 | 1350,13 | 250240,0 | 1715,00 |

Ковариационный анализ представляет собой метод, соединяющий черты дисперсионного и регрессионного анализа. Применим ковариационный анализ к изучению данных табл.9. В результате анализа данных табл.9 мы хотим узнать, зависит ли цена дома от категории, в которую дом был отнесен. Очевидно, что имеется существенное различие между средними ценами по столбцам, дисперсионный анализ наблюдений, несомненно, подтвердит этот вывод. Однако, как нетрудно заметить, дома, считающиеся неплохими, в таблице 9 имеют сравнительно небольшую полезную площадь, площадь домов, считающихся хорошими, заметно больше. Таким образом, величина , которую часто называют сопутствующей переменной, на самом деле, может иметь даже большее значение для объяснения различий в ценах, чем категория дома. В данном случае сопутствующая переменная является мешающей для оценки значимости влияния категории на цену дома.

Можно предложить следующую модель:

 (1)

Здесь каждое наблюдение представлено в виде суммы среднего по столбцу  (представляющего собой влияние категории, в которую отнесен дом), эффекта, вызванного тем, что соответствующая сопутствующая переменная  отличается от общего среднего сопутствующей переменной  и случайного слагаемого, моделирующего влияние неучтенных факторов.

Надо провести дисперсионный анализ, пользуясь ценами, скорректированными с учетом влияния полезной площади. Скорректированные цены определяются следующим образом:

 (2)

Если коэффициент  положителен, то есть  и положительно связаны, тогда, если , то скорректированное значение  окажется меньше, чем исходное значение . В случае отрицательного  и выполнения неравенства  результат коррекции окажется противоположным. Значение оценки  дает регрессионный анализ. Затем, после учета коррекции, проводится дисперсионный анализ скорректированных данных. В модели (1) предполагается, что коэффициент  один и тот же для всех классов. Иначе говоря, влияние полезной площади на цену дома одно и то же независимо от категории. Это предположение об отсутствии взаимодействия между сопутствующей переменной и схемой классификации существенно для последующего анализа. Вообще говоря, прежде чем приступать к ковариационному анализу, надо проверить справедливость предположения об общем коэффициенте. Проверка заключается в подборе линии регрессии для каждого из классов вида:



Затем проверяется гипотеза, что все  равны.

***Задание 1. Проверить гипотезу***

Альтернативная гипотеза:

не все коэффициенты равны.

Если эта гипотеза принимается, то рассматривается общий угловой коэффициент и его оценка .

Введем обозначения:

, 

  

Пользуясь методом наименьших квадратов, найдем объединенную оценку :



Остаток наименьших квадратов для (1) равен:



Оценкой для является скорректированное среднее , поэтому

.

Скорректированная сумма квадратов остатков  равна:



Величина называется скорректированной остаточной вариацией.

Величина имеет  степеней свободы, а объясненная вариация – одну степень свободы, то  имеет  степеней свободы, где 

***Задание 2. Проверить гипотезу*** 

Альтернативная гипотеза:

не все математические ожидания по столбцам равны.

Если нулевая гипотеза верна, то модель (1) принимает вид:

, (3)

оценки метода наименьших квадратов для  будут:



В этом случае остатки для модели (3) будут вычисляться по формуле:



а сумма квадратов остатков будет равна:



где  называется скорректированной остаточной вариацией при условии, что справедлива нулевая гипотеза. Так как имеет степень свободы, то  имеет 

степеней свободы.

Расхождение между  и  должно быть мало, если верна нулевая гипотеза. Обозначим скорректированную вариацию столбцов .

Тогда отношение



Подчиняется распределению Фишера со степенями свободы .

Оценкой для дисперсии служит

.

**II. Линейные контрасты.**

Если нулевая гипотеза о равенстве математических ожиданий отвергается, то может представлять интерес вопрос о том, какие именно средние были причиной этого.

Линейный контраст имеет вид:

, .

Оценка для контраста



Можно показать, что оценка вариации будет равна:



Границы для доверительного интервала будут иметь вид:



Нулевая гипотеза для контраста принимается, если ноль содержится в доверительном интервале

(речь идет о гипотезе )

**III. Задание.**

1. Проведите ковариационный анализ данных в таблице 9.
2. Найдите в интернете данные социально-экономического характера для решения нескольких задач по теме.
3. Сделайте выводы экономического характера.
4. Подготовьте отчет.