### **Множественные сравнения средних**

Результат дисперсионного анализа, указывающий, что средние значения отклика для разных уровней фактора, различаются, не является окончательным результатом анализа изучаемого явления. Это скорее промежуточный результат, который подразумевает дальнейшее раскрытие того, для каких уровней фактора средние значения больше, для каких меньше, а для каких одинаковы. Основная процедура дисперсионного анализа не дает возможности ответить на эти вопросы.

Самый очевидный и простой вариант решения данной задачи - провести серию попарных сравнений при помощи t-критерия, используя в качестве оценки дисперсии величину  - оценку внутригрупповой дисперсии, полученную в ходе дисперсионного анализа. Такой подход реализуется в так называемом ***методе наименьшей значимой разности*** (LSD). Статистика критерия LSD для проверки гипотезы равенства средних и имеет вид:

.

Если наблюдаемое значение статистики , где  - критическая точка распределения Стьюдента уровня  (или квантиль уровня ) с числом степеней свободы , то нулевая гипотеза отклоняется и принимается гипотеза .

Однако, такой подход является не совсем корректным. Если задать, скажем, 5% уровень значимости, то при каждом сравнении вероятность отклонить нулевую гипотезу будет равна 5%, а при серии попарных сравнений, вероятность отклонить хотя бы одну нулевую гипотезу в таком случае существенно превысит 5%. Например, при попарном сравнении средних 4 групп, эта вероятность составит 26,5 %.

Существуют разные подходы к решению данной проблемы. Один из них – уменьшить уровень значимости при попарном сравнении так, чтобы вероятность хотя бы одного отклонения нулевой гипотезы равнялось заданному уровню значимости. Такой подход реализуется в ***методе Бонферрони*** (правильнее говорить о принципе Бонферрони) множественных сравнений, в котором при каждом попарном сравнении задается уровень значимости , где  - число сравнений. Данная величина гарантирует, что вероятность отклонение нулевой гипотезы (при ее истинности) хотя бы в одном из  сравнений не превзойдет . Действительно, пусть  - вероятность ошибки первого рода при одном сравнении. Тогда вероятность хотя бы раз отклонить нулевую гипотезу в серии из  сравнений равна: . Однако, принцип Бонферрони является чересчур консервативным, он приводит к существенному снижению мощности критерия.

LSD – критерий и критерий Бонферрони занимают как бы самые крайние позиции в ряду критериев множественных сравнений. Среди остальных критериев множественного сравнения средних можно выделить критерии множественных сравнений Шеффе, Ньюмена-Келса, Тьюки и другие.

В ***методе множественных сравнений Шеффе*** для проверки гипотезы равенства средних и  используется статистика:

,

где  – оценка внутригрупповой (остаточной) дисперсии, полученная в ходе дисперсионного анализа. Если наблюдаемое значение статистики , где  - критическая точка распределения Фишера уровня  (или квантиль уровня ) с числом степеней свободы  и , то нулевая гипотеза отклоняется и принимается гипотеза .

Критерий Шеффе также относится к достаточно консервативным критериям, то есть обладает малой мощностью. Более мощными, соответственно, более чувствительными являются критерии Тьюки и Ньюмена-Келса.

В ***методе множественных сравнений Тьюки*** (или достоверно значимой разности – HSD) для проверки гипотезы  против альтернативы  используется статистика:

,

значения которой сравниваются с критическими точками уровня  распределения стьюдентизированного размаха  и  степенями свободы. Если наблюдаемое значение статистики , где  - критическая точка распределения стьюдентизированного размаха уровня  (или квантиль уровня ) с числом степеней свободы  и , то нулевая гипотеза отклоняется и принимается гипотеза .

Заметим, что распределение стьюдентизованного размаха с  и  степенями свободы определяется следующим образом. Пусть  – независимые случайные величины с распределением , а  – их размах. Пусть , такая статистика, что величина , имеет распределение хи-квадрат с  степенями свободы. Тогда распределение величины  называется распределением стъюдентизованного размаха с  и  степенями свободы.

В ***критерии Ньюмана-Келса*** используется та же статистика, что и в критерии Тьюки, однако по другому определяются критические точки. В качестве критических точек критерия Ньюмана-Келса, используются критические точки распределения стьюдентизированного размаха с  и  степенями свободы, где  - число средних, расположенных между  и  в вариационном ряду выборочных средних, включая  и . Например, если сравниваются значения  и  вариационного (упорядоченного) ряда средних, то , если сравниваются значения  и , то  и так далее. В пакете STATISTICA используется модифицированный вариант критерия Ньюмана-Келса, в котором в качестве статистики критерия используется величина

.

Аналогичная статистика используется и в ***критерии Дункана***, но в качестве критических точек берутся точки D-распределения Дункана c  и  степенями свободы, где  - число средних расположенных между  и  в вариационном ряду выборочных средних, включая  и .

Методы множественного сравнения средних можно использовать не только для проверки гипотез о попарном различии средних, а также для проверки гипотез о различии средних для любых выбранных наборов групп. В силу этого, основная гипотеза в данных методах в общем случае имеет вид: , где  некоторые заданные константы, удовлетворяющие условию . Например, при ,, мы будем проверять гипотезу  или . При , , будем проверять гипотезу , то есть, гипотезу однородности первой и совокупности второй и третьей групп и т.д. Линейные комбинации вида: , , то есть величины, пропорциональные разности между средними от средних, называются контрастами.

Критерии LSD, Шеффе, HSD Тьюки легко модифицировать под проверку гипотезы . Например, статистика LSD критерия для проверки гипотезы  будет иметь вид:

.

Критическими точками статистики, по прежнему, будут являться квантили распределения Стьюдента уровня  с числом степеней свободы .

PS:

В пакете STATISTICA в модуле ANOVA в качестве оценок дисперсий выборочных средних используются величины , а для построения ДИ для средних используется статистика . Заметим, что даже если доверительные интервалы для двух средних перекрываются, значение статистики LSD могут быть значимыми. Если же интервалы не перекрываются, значения статистики LSD не могут быть незначимыми!