

Дисперсионный анализ ANOVA позволяет ответить на вопросы – влияет ли номинальный многоуровневый фактор на количественный отклик? Влияет ли количественный фактор на номинальный многоуровневый отклик?

С точки зрения математики anova проверяет равенство средних значений количественной шкалы в нескольких подгруппах.

Требования к исходным данным:

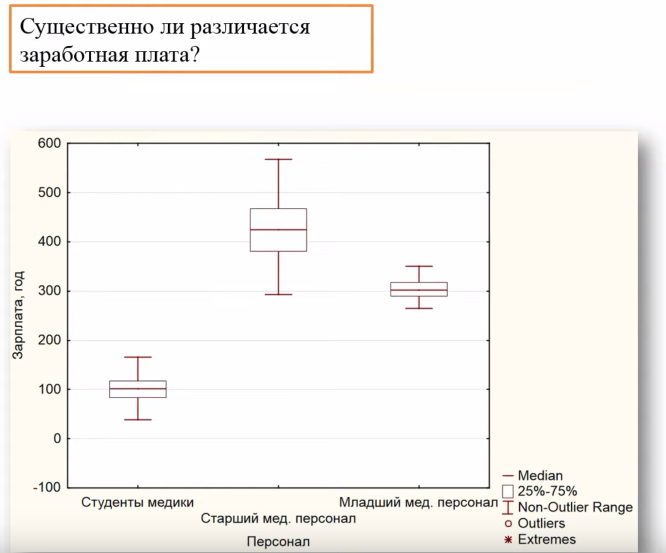
+ номинальная шкала должна иметь 3 и более уровней

+ данные внутри подгрупп, которые образуют эти уровни, должны быть нормально распределены

+ каждый уровень должен иметь хотя бы 30 значений (не строго, главное чтобы было нормально распределено)

В качестве графического представления используется диаграмма ящики с усами.

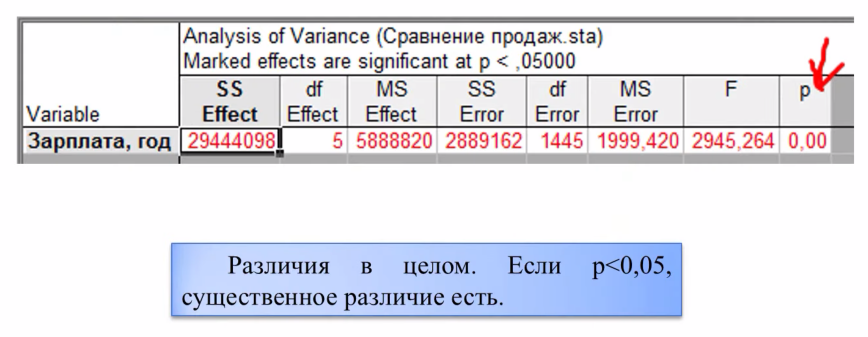
Пример:



Анализируется годовая зп трех трех категорий (студенты медики, старший мед персонал и младший мед персонал) персонала мед учреждения.

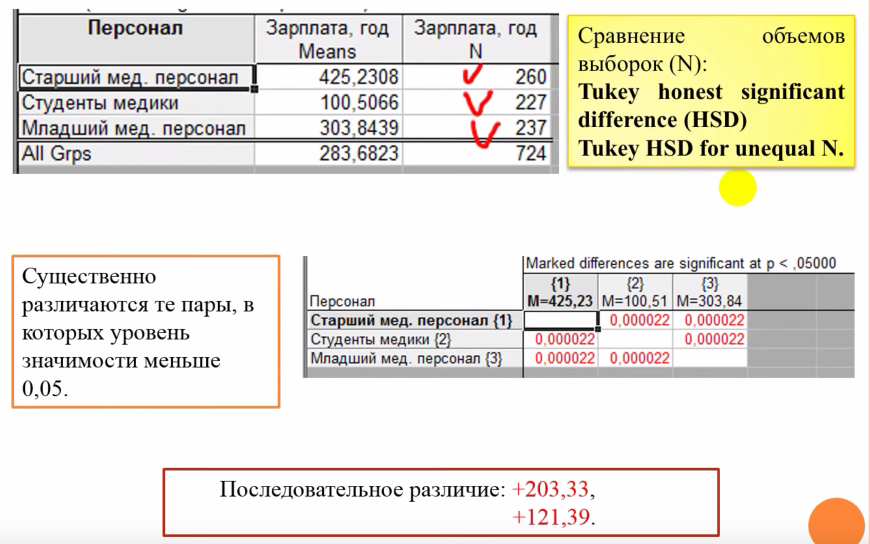
Цель: выяснить, существенно ли различаются в данных категориях зп.

На диаграмме видно, что данные категорий распределены нормально (на самом деле есть небольшая, еле видная, ассиметрия, показать насколько она выражена, покажет аналитическая проверка, кроме того она поможет попарно сравнить и выяснить – например, существенно ли различаются зп старшего и младшего мед персонала).



Сначала проверяется различие в целом по всем группам..

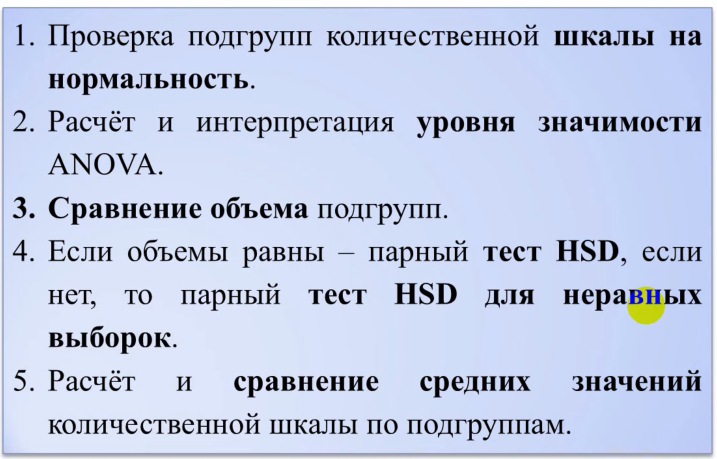
Уровень значимости (p) равен нулю, поскольку он равен нулю, различие существенно.



Затем, чтобы выбрать метод проверки парных различий сравнивается обьемы выборок в подгруппах (N). В данном случае обьемы по подгруппам не равны (если были бы равны, использовался бы HSD как на рисунке, если нет, как в нашем случае, то используется скорректированный HSD (второй в желтом квадрате). Результатом его применения является таблица пониже желтого прямоугольника и он показывает уровень значимости.

После этого сравниваются средние значения и делаем общий вывод.

Алгоритм метода



1. Проверка подгрупп количественной шкалы на нормальность. (anova исполняется только с нормально распределенными данными)
2. Расчет и интерпретация уровня значимости anova.
3. Сревнение обьема подгрупп (количества данных)
4. Если обьемы равны, то используем парный тест hsd, если нет, то парный тест для неравных выборок.
5. Расчет и сравнение средних значений количественной шкалы по подгруппам