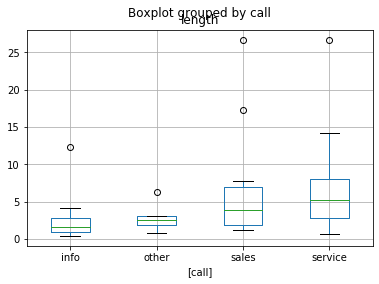
**Задание 1.**

1. **ввести данные в Python:**

df = pd.read\_csv("anova\_lab1.csv")

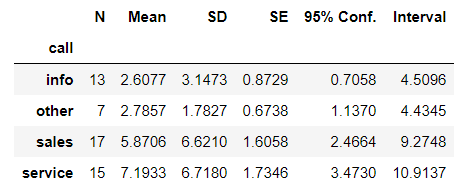
1. **построить в одном масштабе блочные диаграммы для этих четырех видов звонков и описать полученную структуру.**



На полученной диаграмме видно, что звонки типа «инфо» и «другие» длились в основном (за исключением некоторых выбросов) 1-3 минуты.

Звонки типа «продажи» длились уже дольше – до 7 минут, а «обслуживание» - дольше остальных (в среднем) – до 14 минут.

1. **вычислить описательные статистики для каждого вида телефонных звонков**

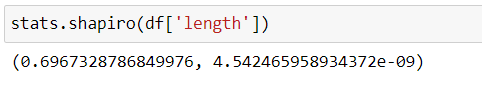


1. **какой из типов звонков имеет наибольшую среднюю продолжительность? какой – наименьшую?**

Наибольшую – service (Столбец Mean – мат. ожидание (среднее значение))

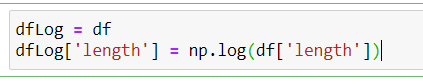
Наименьшую – info

**5) выполняется ли для данного набора данных необходимое для однофакторного дисперсионного анализа предположение о нормальном распределении и равной вариации? почему?**

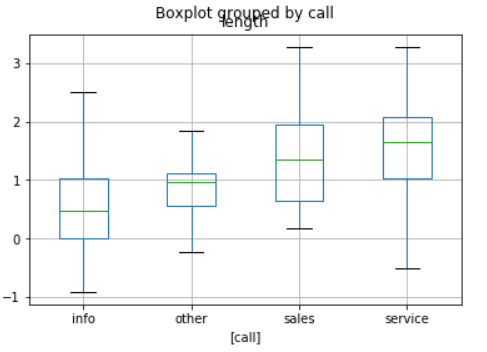


Так как второе значение - p < 0.05 – то гипотезу H0 о нормальном распределении отвергаем (Другими словами – предположение о нормальном распределении и равной вариации НЕ выполняется для данного набора данных)

**6) скопируйте данные под новым именем и вычислите для каждого значения натуральный логарифм.**



**7) Постройте блочные диаграммы для этих логарифмов.**

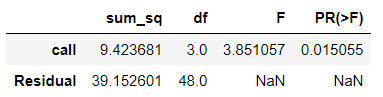


**Можно ли сказать, что предположение о равной вариации лучше выполняется для логарифмов, чем для исходных данных?**

Да, здесь вариация у всех четырех групп в большей степени «равная», чем в исходных данных, где вариация у групп info и other намного меньше, чем у, например, service

**Задание 2.**

**1) Используя значения логарифмов продолжительности телефонных звонков из задания 1, проведите однофакторный дисперсионный анализ:**



sum\_sq - Сумма квадратов

df - степени свободы

F – F-статистика

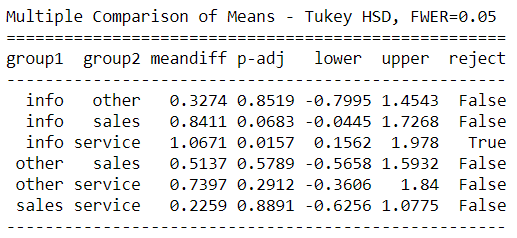
PR(>F) – P-значение

Residual – остаточное значение

**2) опишите полученные результаты о различиях между этими четырьмя видами телефонных звонков;**

В общем случае, если p-значение меньше 0,05, то нулевая гипотеза отвергается (альтернативная гипотеза принимается). Если нулевая гипотеза отвергнута, то можно сделать вывод, что средние значения (мат. ожидания) всех групп не равны.

**3) Проведите тест наименьшего значимого различия (Tukey’s Honestly Significant Difference Test):**



Какая пара типов телефонных звонков существенно отличается друг от друга (в терминах среднего значения логарифма продолжительности звонка)?

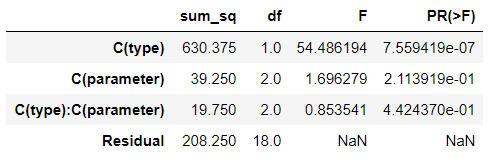
Info и service (так как значение meandiff (дословно – среднее отличие, разница) – самое большое)

**Задание 3.**

1. **Ввести данные в Python:**

df = pd.read\_csv("anova\_lab2.csv")

1. **Проведите двухфакторный дисперсионный анализ.**



1. **Среднее значение эффективности у рекламы 1 выше, чем у рекламы 2. Значимо ли оно выше? Как вы это определили?**

Да, поскольку F = 54.5 (В строке type). Это значение выше табличного (для данной степени свободы – 4.41).

Оно значимо выше, поскольку значение F сильно превышает табличное значение (Или, что аналогично, P-значение крайне низко и намного ниже порогового значения 0.05).

1. **Является ли взаимодействие факторов существенным? О чем это говорит?**

Fтабл = 3.55. Поскольку F < Fтабл, то взаимодействие факторов считаем несущественным. Это говорит о том, что не имеет значения, как эти факторы сочетаются, важно то, что факторы по отдельности сами по себе оказывают влияние на оценку эффективности рекламы.

1. **Является ли воздействие главного эффекта вида рекламы существенным? О чем это говорит?**

Да, аналогично пункту (3). Существенное влияние, по сути, оказывает только типа рекламы.

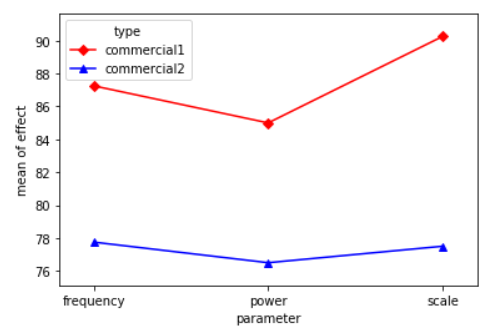
Соответственно, оценка эффективности рекламы зависит от типа рекламы.

1. **Является ли воздействие главного эффекта параметров рекламы существенным? О чем это говорит?**

Воздействие параметров рекламы не являются существенным, так как 1.7 < 3.55 (F < Fтабл).

Это говорит о том, что параметры рекламы не оказывают влияния на оценку эффективности.

1. **Постройте график средних значений для главных эффектов. Сделайте выводы.**



На графике чётко видно, что красная линия, обозначающая первый тип рекламы явно выше синей линии, что, опять-таки, подтверждает существенное влияние типа рекламы.

Можно также увидеть, что параметр рекламы, возможно, влияет на оценку эффективности рекламы (Так как обе линии имеют похожий характер роста и падения), но если это влияние и существует, то оно незначительно (Возможно, необходимо большее число наблюдений, чтобы влияние было более заметно)

**Задание 4.**

В таблице файла Лаб\_Дисп3 приведены результаты исследования Kirby and Dardis, изучавшего на протяжении 13 недель цены 20 наименований товаров. В результате исследования выявлено, что цены в универсальных магазинах действительно на 40% выше цен в магазинах производителя.

1. **Действительно ли высокие цены (в среднем на 40% выше в универмагах) значимо выше? Почему?**

Да, потому что P-значение стремится к 0, а если P < 0.05 – мы отвергаем гипотезу H0, то есть считаем, что магазин оказывает влияние на цену, а так как это значение практически равно нулю, то это влияние – сильное, а значит цены значимо выше.

1. **Какой вид дисперсионного анализа здесь использован?**

Трёхфакторный, потому что исследуются три фактора – тип магазина, вид товара и неделя и их взаимодействия друг с другом.

1. **Определите три фактора в этом анализе. Сколько категорий имеет каждый из факторов?**

Для этого смотрим на степени свободы df.

Количество факторов = df + 1

Таким образом:

Тип магазина – 2 категории

Вид товара – 20 категорий

Неделя – 13 категорий

1. **О чем свидетельствует p-значение для главного эффекта фактора В?**

Аналогично пункту (1) – о том, что вид товара, очевидно, влияет существенно на цену.

1. **Значимо ли различаются цены разных недель? Как это узнать?**

Да, различаются, да, значимо, но не так, как различаются цены разных товаров или в разных магазинах. Узнать это можно из значения F = 12.83, которое превышает Fтабл, которое примерно равно 1.76.

1. **Рассмотрите взаимодействие типа магазина и вида товара. Значимо ли оно? О чем это говорит?**

Да, значимо, из тех же соображений, что и в предыдущих пунктах. Это говорит о том, что следует учитывать, видимо, на чём специализируется магазин (Например, спорттовары и спортмагазин, универсальный магазин и спорттовары, в общем какое-то влияние это будет оказывать)

1. **Рассмотрите взаимодействие типа магазина и недели? Значимо ли оно? О чем это говорит?**

Здесь также видно, что взаимодействие влияет, но гораздо меньше (менее значимо), чем между типом магазина и видом товара. Это может много о чём говорить, но, скорее всего дело в сезоне и в том, как магазин востребован в разные сезоны и отсюда - ценообразование.

1. **Рассмотрите взаимодействие вида товара и недели. Значимо ли оно? О чем это говорит?**

Здесь уже видно, что взаимодействие не оказывает влияния на цену (так как p > 0.05, причём существенно). Следовательно, цена товара не будет зависеть от взаимодействия факторов вида товара и недели.

1. **Обычно мы изучаем p-значения, чтобы посмотреть, достаточно ли они малы, чтобы можно было заявить, что результат является статистически значимым. Однако для трехфакторного взаимодействия p-значение подозрительно велико, что наводит на мысль о том, что здесь значительно *меньше* случайности, чем можно было бы ожидать для этой модели. Какое техническое предположение для проверки гипотезы может быть здесь не выполнено?**

Не выполняется предположение об отсутствии мультиколлинеарности между независимыми переменными.