## Задача

Проверить воспроизводимость результатов оценки когерентности ЭЭГ при помощи непараметрических тестов

Рассчитывались оценки функции когерентности ЭЭГ одних и тех же испытуемых в состояниях покоя с открытыми и закрытыми глазами, всего 177 испытуемых.

Для всех 177 испытуемых были посчитаны значимые каналы. Однако при уменьшении выборки уменьшалось и количество значимых каналов. Для установки значимых каналов использовались параметрические тесты. В качестве альтернативы было предложено использовать непараметрические тесты, в частности, перемешивания.

В предыдущем анализе разбиение на более мелкие группы проводилось вручную. Чтобы получить более общую картину воспроизводимости паттернов, предложено провести многократное рандомизированное разбиение на более мелкие группы.

## Методы

**Непараметрические тесты**

В качестве альтернативы параметрическим тестам (которые имеют некоторые ограничения) используются непараметрические тесты, например, nonparametric permutation testing. Permutation test более общий, так как не основывается на предположении о распределении данных, в сравнении с параметрическими тестами. При использовании перестановок, мы предполагаем, основываясь на данных, как бы выглядело распределение тестовой статистики, если бы нулевая гипотеза была верна. [[1]](https://www.zotero.org/google-docs/?uF3RVc)

Процедура тестирования описана в статье [[2]](https://www.zotero.org/google-docs/?7gulCt). Нулевая гипотеза теста на перестановки – при перемешивании двух состояний А и В не будет значительной разницы в тестовой статистике до перемешивания и после. Цель тестирования с помощью перемешиваний – определить вероятность того, что наблюдаемая тестовая статистика могла быть получена при верной нулевой гипотезе. Альтернативная гипотеза: наблюдаемая тестовая статистика не случайна, обнаружена не из-за выбросов, а эффект действительно присутствует.

Чтобы получить распределение статистической величины при верной нулевой гипотезе, мы должны повторить процедуру много раз. Далее сравниваем эмпирическую тестовую статистику со статистикой, полученной при перемешивании. Если эмпирическая статистика лежит внутри распределения при верной нулевой гипотезе, нулевая гипотеза не может быть отвергнута. Это может значить, что разницы между двумя состояниями нет, а метки испытаний А и В могли быть получены случайным образом.

Если же эмпирическая статистика далека от распределения при верной нулевой гипотезе, нулевая гипотеза отвергается, а эффект можно считать статистически значимым.

В нашем случае, для каждого ритма и для каждой пары каналов считается разница *открытые – закрытые* глаза. Далее для каждого человека случайно путаем состояние: *закрытые – открытые* глаза, считаем разницу, и вычисляем среднее для пары каналов. Если эффекта нет, при таком перемешивании мы не увидим разницу между эмпирическим средним и полученным с помощью перемешиваний.

Далее нужно вычислить p-value. Для этого мы считаем, сколько раз значения статистики при нулевой гипотезе (среднее после перемешивания) получились более экстремальными, чем значение эмпирической статистики, и делим на количество итераций перемешивания. Чем больше итераций, тем точнее p-value.

**Cluster-based permutation testing**

В статье [[3]](https://www.zotero.org/google-docs/?CKd6Io) предложен алгоритм использования cluster-based permutation test для работы с ЭЭГ и МЭГ. Этот метод популярен, его имплементируют во многих пакетах для обработки нейрофизиологических и нейровизуализационных данных для Matlab и Python. [[3–5]](https://www.zotero.org/google-docs/?FuhTSN)

Метод используется для коррекции множественных сравнений при работе с данными пространственно-временной структуры. Идея cluster-based permutation test состоит в поиске групп электродов и временных точек, в которых наблюдается значимое различие между состояниями. Кластер – это группа точек в пространстве время-частота-электрод, активность которых была сильнее выражена в одном состоянии в сравнении с другим. Цель – найти кластеры, которые можно было бы считать значимыми. Данный метод может быть использован в экспериментах не только с разными состояниями, но и для сравнения групп.

Сперва нужно найти кластеры, в которых наблюдается эффект. Это можно сделать имея предварительные знания о том, где искать эффект. Либо выбрать те электроды в те временные периоды, когда наблюдалось значимое различие между состояниями:

1. Для каждого канала в каждой временной точке считаем разницу между состояниями и проверяем значимость (например, с помощью t-теста, как в оригинальной статье [[3]](https://www.zotero.org/google-docs/?f36P8N), или теста Уилкоксона–Манна–Уитни [[6]](https://www.zotero.org/google-docs/?M02Jwj))
2. Выбираем те точки, значимость которых больше некоторого порога
3. Составляем кластеры из значимых точек, основываясь на их близости в смысле времени или положения электродов
4. Считаем статистику кластера – суммируем значения всех точек кластера
5. Сохраняем кластер с наибольшей суммой – сумма определяет “размер” кластера

Эффект считаем значимым, если он появляется в кластере “достаточного большого” размера. “Достаточный” размер кластера определяем с помощью теста на перемешивание.

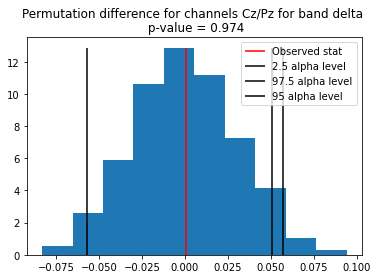
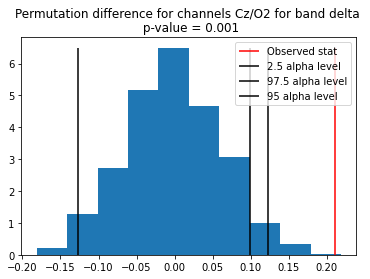
Перемешиваем состояния и делим на 2 выборки того же размера, что и оригинальные выборки, и уже для новых выборок ищем кластеры, в которых наблюдается эффект. Суть теста на перемешивание аналогична описанному в предыдущем разделе. На каждой итерации смотрим на то, какого размера появляются кластеры при случайном перемешивании состояний, сохраняем размер самого большого кластера. В итоге получаем распределение размеров кластеров, которые могут получиться по воле случая. Основываясь на распределении, выбираем порог. Если мы получили кластер большего размера, чем пороговое значение, будем считать такой кластер значимым.

## Демонстрация этапов анализа

**Распределение при перемешивании**, 1000 итераций.

Для Cz/O2 delta, наблюдаемое среднее (красная линия) лежит за пределами распределения, a p-value < 0.001, то есть полученное значение для этой пары не случайно, канал значимый.

Для Cz/Pz delta, наблюдаемое среднее лежит внутри полученного распределения, здесь разницу можно считать незначительной.



**Процедура проверки**

Формируем подвыборку или используем целую. Повторяем много раз:

1. Для каждого канала:
   1. Вычисляем разницу между состояниями, считаем среднее по испытуемым
   2. Вычисляем разницу между перепутанными состояниями, считаем среднее по испытуемым
   3. Вычисляем уровень значимости

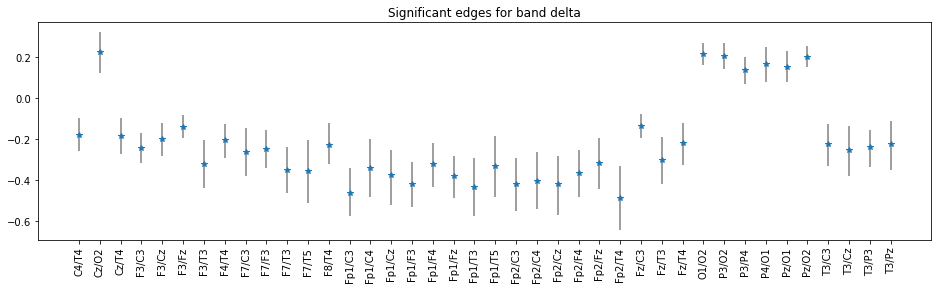
Сколько раз полученное среднее превысило реальное

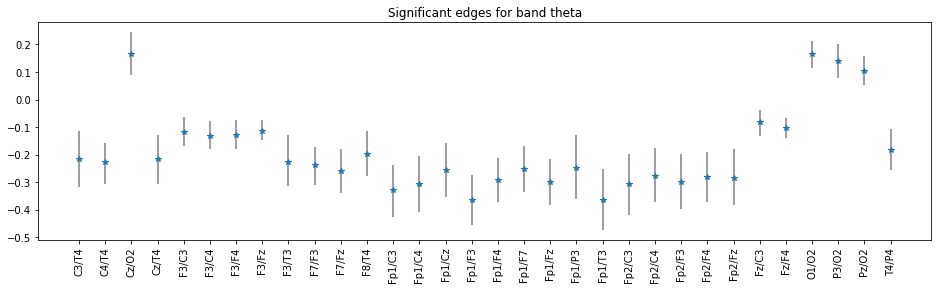
1. Запоминаем значимые пары каналов

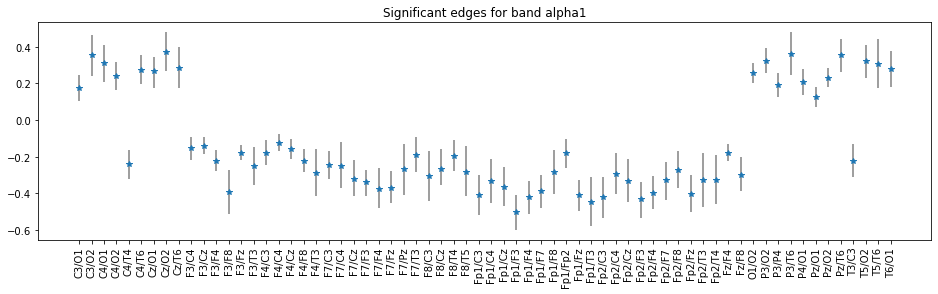
Задаем порог, при котором будем считать пару каналов значимой. Значимые пары – чей p-value меньше порога значимости

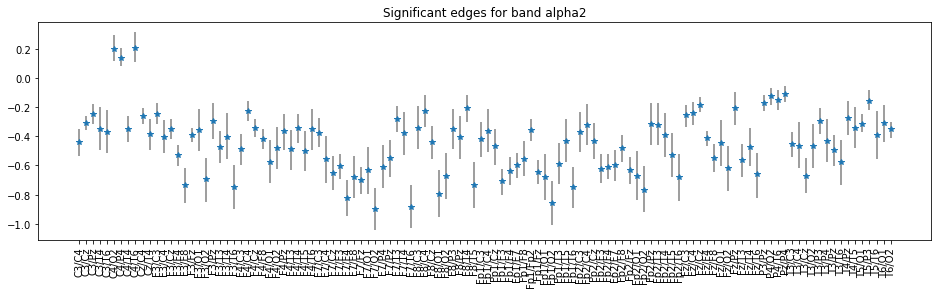
Бутстрап – метод для получения доверительных интервалов. Из выборки случайным образом составляются тысячи новых подвыборок. Для составления подвыборки, элементы из выборки выбираются случайно с повторениями (тогда один испытуемый может попасть в подвыборку несколько раз) или без. Далее считаем статистику для подвыборок (среднее в нашем случае), и, таким образом, можем получить доверительные интервалы для средней разницы состояний.

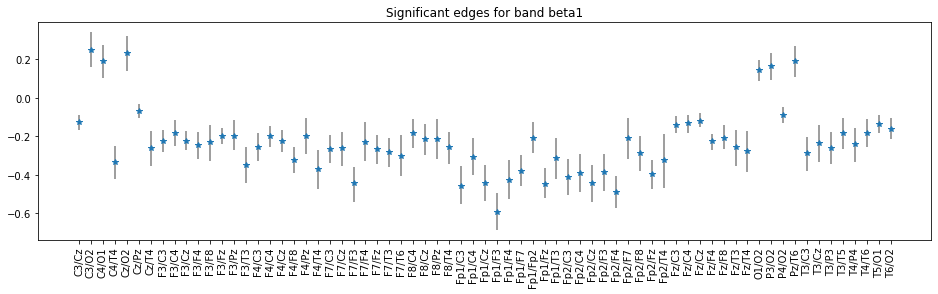
Для конкретного эксперимента с группой размера 100 нарисуем доверительные интервалы для значимых пар каналов с помощью бутстрапа с повторениями.

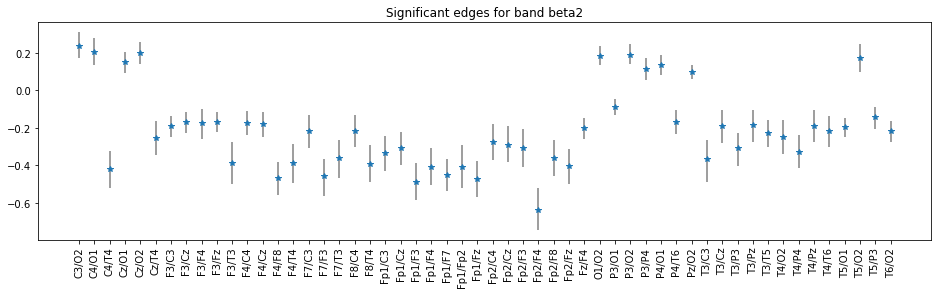


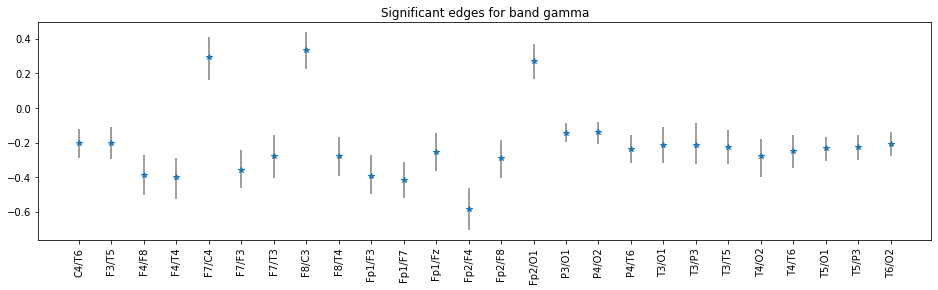










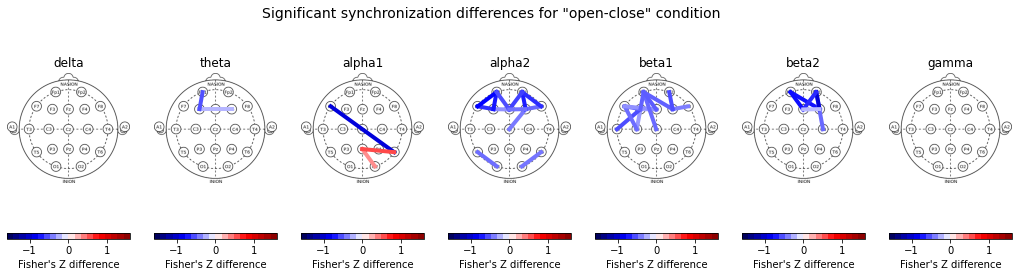


**Визуализация значимых пар каналов для одного эксперимента**

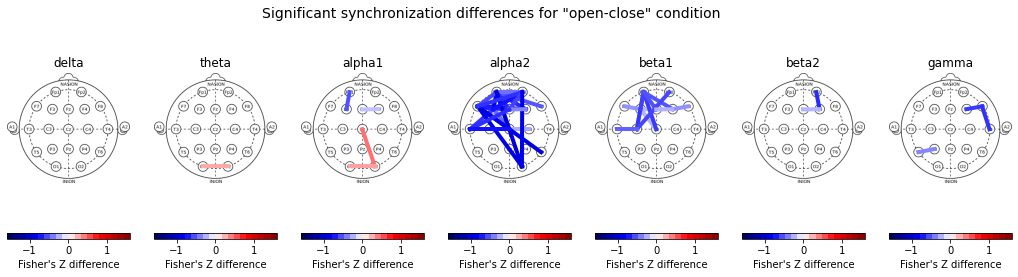
Получим значимые пары каналов для конкретного эксперимента с помощью теста на перемешивания. Число симуляций = 10 000, уровень значимости с поправкой на множественные сравнения = 0.05 / 171.

Посмотрим как будет результат меняться для разного размера выборок. При каждом новом запуске результаты будут отличаться, так как разбиение на группы происходит случайно.

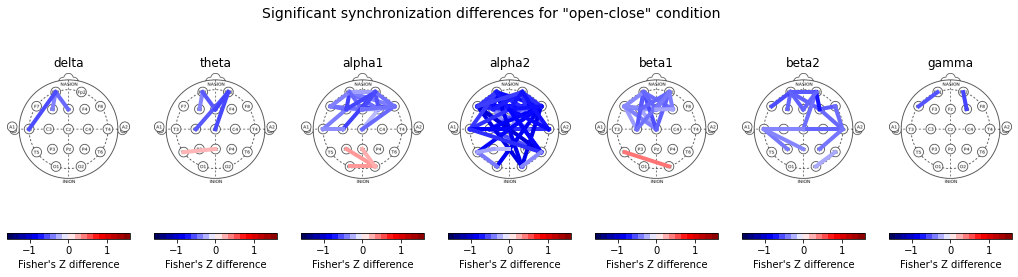
размер группы = 30



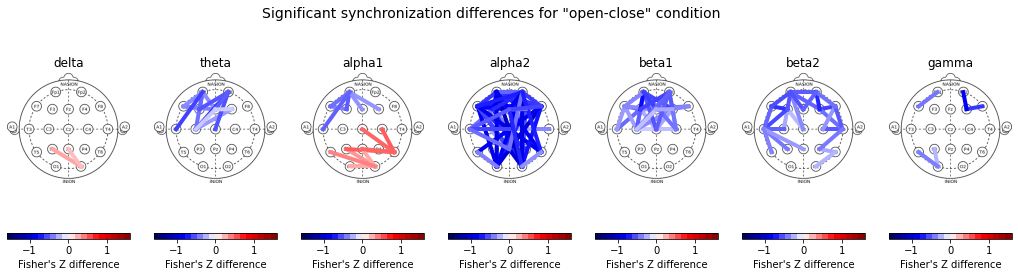
размер группы = 40



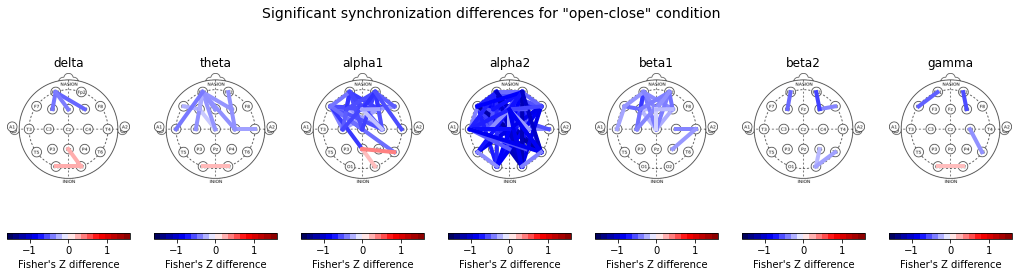
размер группы = 50



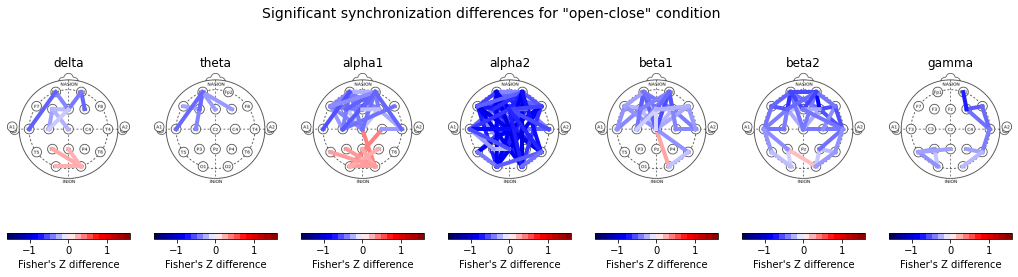
размер группы = 60



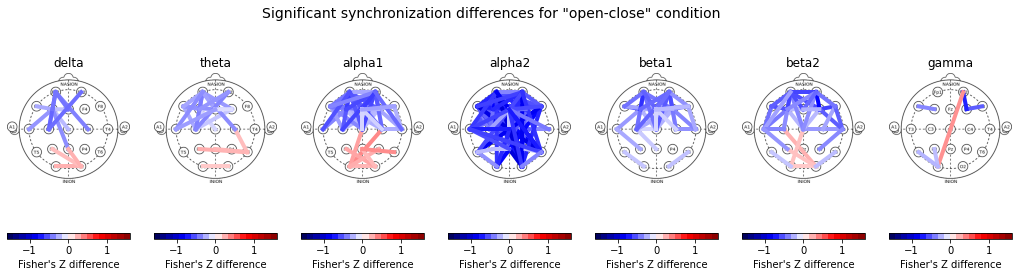
размер группы = 70



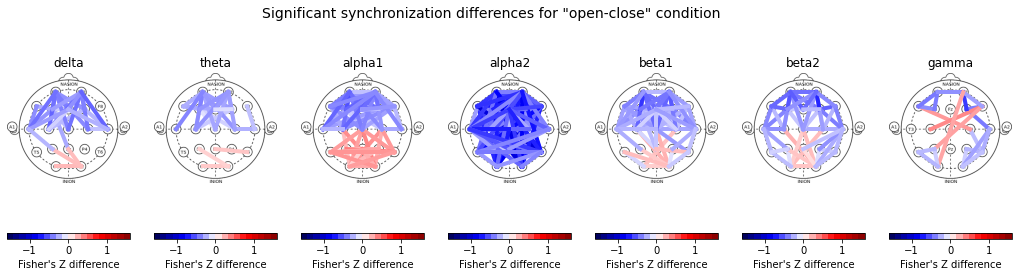
размер группы = 80



размер группы = 100

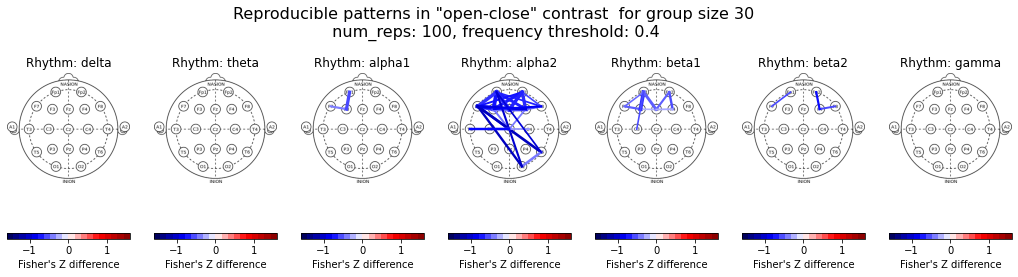


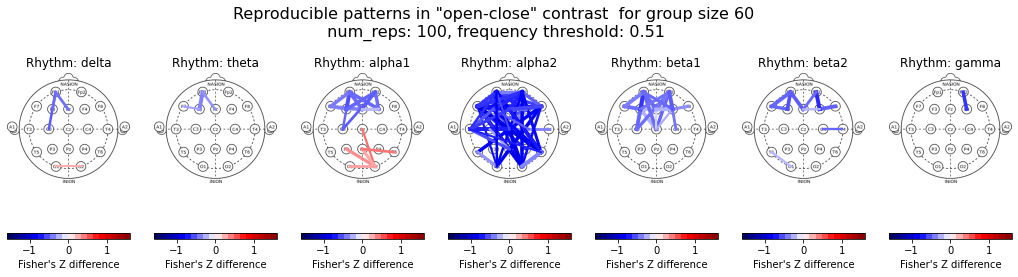
размер группы = 177

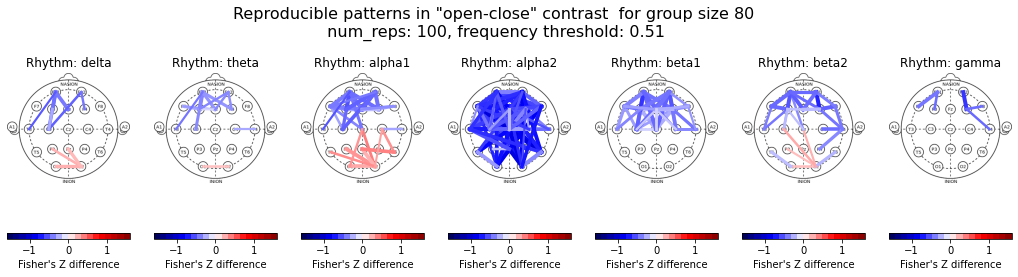


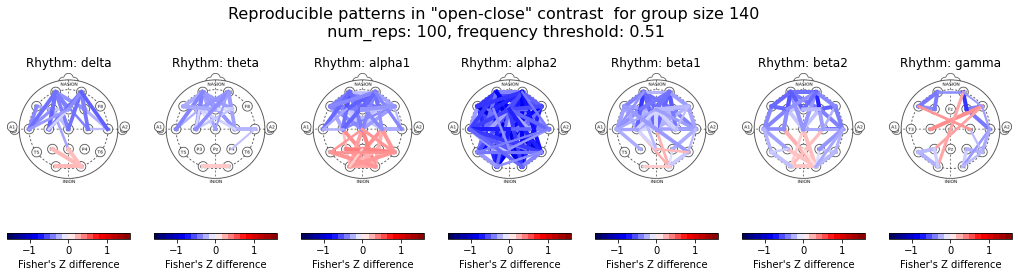
## Оценка воспроизводимости

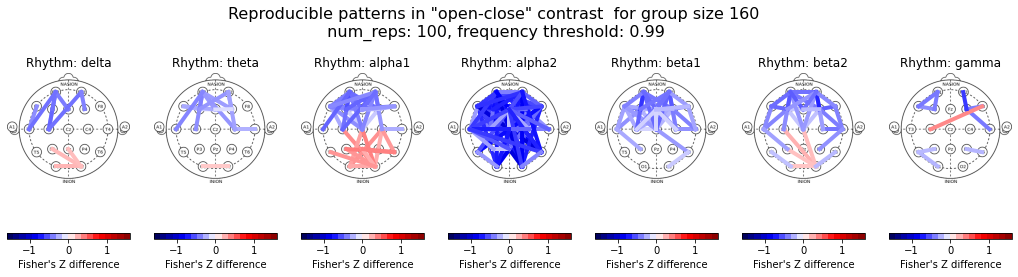
Посчитаем те же самые значимые разности, но добавим условие воспроизводимости. Из целой выборки набираем подвыборку и запоминаем ее значимые каналы – это один эксперимент. Проведем много таких экспериментов и из всех полученных значимых каналов оставим только те, что воспроизводятся чаще, чем в некоторой доле случаев (например, в половине экспериментов). При этом порог для уровня значимости предлагается увеличить, так как в данном случае мы накладываем дополнительные ограничения, и условие на множественные сравнения можно ослабить.

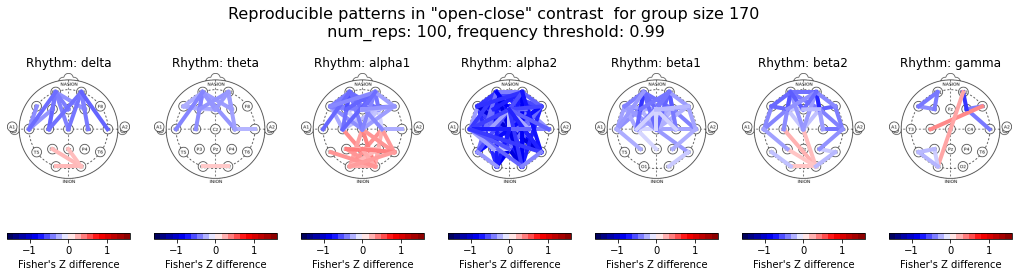












Таблицы со значимыми каналами:

<https://drive.google.com/drive/folders/1YXL1rRJvJslXW8CoilG1xrP-A4MaxJQF?usp=sharing>

## Выводы

* при многократном повторении эксперимента с одним и тем же размером группы результаты могут отличаться очень существенно.
* при росте испытуемых наблюдается увеличение количества значимых “закономерностей”, которые не обнаруживаются на меньшем объеме выборки
* большой объем выборки дает возможность для увеличения порога воспроизводимости. Для группы размера 170 испытуемых и порога 0.99 получаем близкую картину паттернов с меньшим размером группы.

**Вопросы и предложения:**

* как вычислялась синхронизация? в выборке в основном отрицательные значения, при этом если смотреть стандартный подход к вычислению фазовой или спектральной синхронизации [1] - значения до преобразования фишера лежат в диапазоне 0-1 и не могут быть отрицательными
* можно смотреть воспроизводимость для разных размеров группы и выделять только те паттерны которые стабильны на разных размерах выборок.
* можно вычислять коэффициенты воспроизводимости, но в данном контексте они не очень информативны, так как теряется информация о конкретных парах. При этом можно пробовать выделять какие-то конкретные паттерны, например как для дельта ритма видимо связку внизу из трех пар каналов O2-O1, O2-Pz, O2-P3 и пытаться как то их интерпретировать

## Литература

[1. Cohen M.X. Analyzing Neural Time Series Data: Theory and Practice. Analyzing Neural Time Series Data. – The MIT Press, 2014.](https://www.zotero.org/google-docs/?XSysdG)

[2. Maris E., Schoffelen J.-M., Fries P. Nonparametric statistical testing of coherence differences//Journal of Neuroscience Methods,  2007,  Vol. 163,  No. 1,  P. 161-175.](https://www.zotero.org/google-docs/?XSysdG)

[3. Maris E., Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data//Journal of Neuroscience Methods,  2007,  Vol. 164,  No. 1,  P. 177-190.](https://www.zotero.org/google-docs/?XSysdG)

[4. Puoliväli T., Palva S., Palva J.M. Influence of multiple hypothesis testing on reproducibility in neuroimaging research: A simulation study and Python-based software//Journal of Neuroscience Methods,  2020,  Vol. 337,  Influence of multiple hypothesis testing on reproducibility in neuroimaging research,  P. 108654.](https://www.zotero.org/google-docs/?XSysdG)

[5. Esch L., Dinh C., Larson E., Engemann D., Jas M., Khan S., Gramfort A., Hämäläinen M.S. MNE: Software for Acquiring, Processing,and Visualizing MEG/EEG Data//Magnetoencephalography: From Signals to Dynamic Cortical Networks/ eds. S. Supek, C.J. Aine. – Cham: Springer International Publishing, 2019. – MNE. – P. 1-17.](https://www.zotero.org/google-docs/?XSysdG)

[6. Candia-Rivera D., Valenza G. Cluster permutation analysis for EEG series based on non-parametric Wilcoxon–Mann–Whitney statistical tests//SoftwareX,  2022,  Vol. 19,  P. 101170.](https://www.zotero.org/google-docs/?XSysdG)